



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 105953

TO: Karen A Lacourciere
Location: CM-1/11D09/11E12
Art Unit: 1635
Wednesday, October 15, 2003

Case Serial Number: 09/915543

From: Edward Hart
Location: Biotech-Chem Library
CM1-6B02
Phone: 305-9203

edward.hart@uspto.gov

Search Notes

Examiner Lacourciere,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

STIC-Biotech/ChemLib

105953

From: Lacourciere, Karen
Sent: Tuesday, October 14, 2003 3:13 PM
To: STIC-Biotech/ChemLib
Subject: Sequence search request 09/915,543

Please perform the following sequence searches for 09/915,543 in the commercial databases, pre-grant pubs and pending files (interference)

- 1) Residues 177-204 of SEQ ID NO:15
- 2) Residues 349-383 of SEQ ID NO:15
- 3) Residues 199-392 of SEQ ID NO:15

Thank-you!

Karen A. Lacourciere Ph.D.

CM1 11D09 GAU 1635
(703) 308-7523
mailbox 11E12

RECEIVED
OCT 14 2003
STIC

Received from:
Technical Info. Specialist
STIC/Biotech
CAM 0910 Tel 308-9203

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 10/15/03
Date Completed: 10/15/03
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: 3
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: DD
WWW/Internet: _____
Other (specify): _____



STIC SEARCH RESULTS

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*:

Mary Hale, Information Branch Supervisor
308-4258, CM1-1E01

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1 - Circ. Desk



Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions .rnpm and .rnpn

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions .rapm and .rapn

Because they contain data that is confidential, the results of Pending database searches should not be left in the case .

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 17.214 Seconds
(without alignments)
258.182 Million cell updates/sec

Title: US-09-915-543-15_COPY_177_204

Perfect score: 136

Sequence: 1 VYFSTEMAKAEAVLKGQVEIVSFH 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq_19Jun03.*

```

1: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136	100.0	1426	23 AAB71229	Human legless homo
2	136	100.0	1435	22 ABB11808	Human BCL9 homolog
3	129	94.9	1420	22 AAO05855	Human polypeptide
4	98	72.1	1429	22 ABB58779	Drosophila melanog
5	98	72.1	1464	23 AAB71228	D. melanogaster 19
6	94	69.1	1494	23 AAU78460	Mouse beta-catenin
7	91	66.9	114	23 ABP06595	Human ORFX protein
8	67	49.3	320	23 AAU78461	Mouse beta-catenin
9	65	47.8	113	22 AAO07544	Human polypeptide

10	52	38.2	360	21	AAG33446	Zea mays protein f
11	52	38.2	448	21	AAG33445	Zea mays protein f
12	52	38.2	509	21	AAG33444	Zea mays protein f
13	51	37.5	1049	22	ABB60387	Drosophila melanog
14	50	36.8	130	23	ABB89793	Human polypeptide
15	50	36.8	365	22	ABG15088	Novel human diagno
16	50	36.8	621	23	AAE15740	Human aminoacyl-tr
17	50	36.8	631	21	AAB43285	Human aminoacyl tr
18	50	36.8	1063	23	ABB08919	Bifidobacterium lo
19	49	36.0	187	23	ABP66271	Aspergillus fumiga
20	49	36.0	984	24	ABJ25889	Aspergillus fumiga
21	49	36.0	1058	24	ABJ26489	Aspergillus fumiga
22	48	35.3	498	21	AAE17973	Arabidopsis thalia
23	48	35.3	498	21	AAE17972	Arabidopsis thalia
24	48	35.3	609	21	AAE17972	Arabidopsis thalia
25	48	35.3	609	21	AAE17971	Arabidopsis thalia
26	48	35.3	630	21	AAE17971	Arabidopsis thalia
27	48	35.3	630	21	AAE17971	Arabidopsis thalia
28	47	34.6	886	21	AAE17971	Arabidopsis thalia
29	47	34.6	905	21	AAE17971	Arabidopsis thalia
30	47	34.6	1095	21	AAE17971	Arabidopsis thalia
31	47	34.6	1114	21	AAE17971	Arabidopsis thalia
32	47	34.6	1168	21	AAE17971	Arabidopsis thalia
33	47	34.6	1187	21	AAE17971	Arabidopsis thalia
34	47	34.6	2639	20	AAW73476	Grapevine leafroll
35	45.5	33.5	341	22	ABB70889	Drosophila melanog
36	45	33.1	401	22	ABG04882	Novel human diagno
37	44.5	32.7	119	21	AAE17971	Arabidopsis thalia
38	44.5	32.7	137	21	AAE17971	Arabidopsis thalia
39	44.5	32.7	152	21	AAE17971	Arabidopsis thalia
40	44.5	32.7	160	21	AAE17971	Arabidopsis thalia
41	44.5	32.7	170	21	AAE17971	Arabidopsis thalia
42	44.5	32.7	170	21	AAE17971	Arabidopsis thalia
43	44.5	32.7	177	21	AAE17971	Arabidopsis thalia
44	44.5	32.7	447	21	AAE17971	Arabidopsis thalia
45	44.5	32.7	450	22	AAE17971	S. epidermidis ope

ALIGNMENTS

RESULT 1

AAE171229
ID AAB71229 standard; Protein: 1426 AA.

XX AAB71229;

XX AC AAB71229;

XX DT 18-NOV-2002 (first entry)

XX DE Human legless homologue lgs/bcl9 protein.

XX KW Legless; human; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;
XX KW tissue proliferation; tumour; cytotstatic; cellular disorder; colon;
XX KW blood disorder; cancer; breast; head and neck cancer; brain; thyroid;
XX KW medulloblastoma; skin cancer; tissue regeneration; tissue repair.

XX OS Homo sapiens.

XX PN US2002086986-A1.

XX PD 04-JUL-2002.

XX PF 27-JUL-2001; 2001US-0915543.

XX PR 28-JUL-2000; 2000US-221502P.

XX PR (BASL/) BASLER K.

XX PA (BRUN/) BRUNNER E.

XX PA (FROE/) FROESCH B.

XX PA (KRAM/) KRAMP T.

XX PI (PETE/) PETER O.

PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;

XX WPI; 2002-635689/68.
 DR N-PSDB; AAF88467.
 XX
 XX Novel polypeptide useful in therapeutic method for treating disorders
 PT of cell fate such as cell differentiation or cell proliferation -
 XX
 XX Example II; Fig 8B; 41pp; English.
 XX
 CC This invention describes a novel polypeptide sharing one or more
 CC homologous amino acid domains with the legless (lgs) protein, a
 CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway
 CC involved in the formation and maintenance of spatial arrangements
 CC and proliferation of tissues during development, and in the formation
 CC and growth of many human tumours. The products of the invention have
 CC cytoskeletal activity and can be used to treat cellular disorders, blood
 CC disorders and cancers caused by over-stimulation of the Wnt pathway,
 CC where the cancerous condition is colon, breast, head and neck, brain,
 CC thyroid, medulloblastoma or skin cancer. The product could also be used
 CC to promote tissue regeneration and repair. This sequence represents the
 CC human legless (lgs) protein homologue lgs/bcl9 described in the
 CC disclosure of the invention.
 XX
 SQ Sequence 1426 AA;
 Query Match 100.0%; Score 136; DB 23; Length 1426;
 Best Local Similarity 100.0%; Pred. No. 2.9e-12;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28
 DB 177 VYVFSTEMANKAAEAVLKQGVETIVSFH 204
 RESULT 2
 ID ABB11808
 XX ABB11808 standard; peptide; 1435 AA.
 AC ABB11808;
 XX
 DT 11-JAN-2002 (first entry)
 DE Human BCL9 homologue, SEQ ID NO:2178.
 KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
 KW chronic inflammatory condition; proliferative retinopathy;
 KW atherosclerosis; coronary heart disease; arterial ischaemia;
 KW bone disorder; osteoporosis; vascular growth disorder; immune disorder;
 KW tissue regeneration; wound healing; infection; antiinflammatory;
 KW cell culture; drug screening; gene therapy; antitumor; antitumor;
 KW antasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
 KW cytoskeletal; osteopathic; vasotropic; cardiant; virucide; antibacterial;
 KW antifungal; vulnery; antitumor.
 XX
 OS Homo sapiens.
 XX
 PN WO200157188-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US03800.
 XX
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX

DR WPI; 2001-457740/49.
 DR N-PSDB; ABA09052.
 XX
 XX Human proteins and DNA encoding sequences useful for preventing,
 PT treating or ameliorating a medical condition in a mammalian subject
 PT e.g. arthritis and cancer -
 XX
 XX Claim 20; Page 256-257; 1963pp; English.
 XX
 CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
 CC invention also relates to vectors and recombinant host cells comprising a
 CC nucleotide of the invention, methods of producing the novel polypeptides,
 CC antibodies against the polypeptides, methods of detecting the nucleotides
 CC or polypeptides in a sample, and methods of identifying compounds which
 CC bind to polypeptides of the invention. Although novel, many of the
 CC polypeptides of the invention have homology to known proteins, thereby
 CC giving an insight into their probable biological activities, and hence
 CC potential therapeutic applications. The polypeptides of the invention may
 CC have various activities, including cytokine, cell proliferation or cell
 CC differentiation activities; stem cell growth factor activity;
 CC haematopoiesis regulatory activity; tissue growth activity;
 CC immunomodulatory activity; activin- or inhibin-related activities;
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
 CC thrombolytic activities; receptor or ligand activities; or may be
 CC involved in oncogenesis, cancer cell proliferation or metastasis.
 CC Depending on their biological activities, polypeptides and nucleotides of
 CC the invention are useful for preventing, treating or ameliorating medical
 CC conditions, e.g., by protein or gene therapy. Such conditions include
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
 CC vascular growth. Polypeptides involved with tissue regeneration and
 CC repair (or nucleic acids encoding them) may be used to promote wound
 CC healing (e.g., of burns, incisions and ulcers), while those with
 CC immunomodulatory activities may be used in the treatment of viral,
 CC bacterial and fungal infections in addition to immune disorders.
 CC Polypeptides with growth factor activity may be used in cell cultures to
 CC promote cell growth. For example, such polypeptides may be used to
 CC manipulate stem cells in culture to give rise to neuroepithelial cells
 CC that can be used to augment or replace cells damaged by illness,
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides
 CC may also be used in the diagnosis of the above conditions, and in drug
 CC screening techniques. The present sequence represents a novel human
 CC polypeptide of the invention.
 XX
 SQ Sequence 1435 AA;
 Query Match 100.0%; Score 136; DB 22; Length 1435;
 Best Local Similarity 100.0%; Pred. No. 2.9e-12;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28
 DB 217 VYVFSTEMANKAAEAVLKQGVETIVSFH 244
 RESULT 3
 ID AAO05855 standard; Protein; 140 AA.
 XX
 AC AAO05855;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 19747.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.
 XX

OS Homo sapiens.
 XX WO200164835-A2.
 XX
 XX
 PD 07-SEP-2001.
 XX
 XX 26-FEB-2001; 2001WO-US04927.
 XX
 XX 28-FEB-2000; 2000US-0515126.
 PR 18-MAY-2000; 2000US-0577409.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Drmanac RT;
 PI WPI; 2001-514838/56.
 XX N-PSDB; AAI85786.
 DR
 DR
 XX Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -
 PT
 PS Claim 20; SEQ ID NO 13747; 1399pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 140 AA;
 SQ
 Query Match 94.9%; Score 129; DB 22; Length 140;
 Best Local Similarity 92.9%; Pred. No. 2.3e-12;
 Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYVFSTEMANKAAKAVLKGQVETIVSFH 28
 DB 59 VYVFSTEMANKAAKAVLKGQVETIVSFH 86
 RESULT 4
 ABB58779
 ID ABB58779 standard; Protein; 1429 AA.
 XX
 AC ABB58779;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 DE Drosophila melanogaster polypeptide SEQ ID NO 3129.
 XX
 XX Drosophila: developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 KW
 OS Drosophila melanogaster.
 XX
 XX WO200171042-A2.
 PN
 XX
 PD 27-SEP-2001.
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 PF
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 PR

XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL02882.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PT
 XX Disclosure; SEQ ID NO 3129; 2lpp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABB01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 1429 AA;
 SQ
 Query Match 72.1%; Score 98; DB 22; Length 1429;
 Best Local Similarity 57.1%; Pred. No. 2.8e-06;
 Matches 16; Conservative 9; Mismatches 3; Indels 0; Gaps 0;
 QY 1 VYVFSTEMANKAAKAVLKGQVETIVSFH 28
 DB 323 IFVFTQLANKGAESVLSGQFTLIAYH 350
 RESULT 5
 AAB71228
 ID AAB71228 standard; Protein; 1464 AA.
 XX
 AC AAB71228;
 XX
 XX 18-NOV-2002 (first entry)
 DT
 DE D. melanogaster lgs protein.
 XX
 XX Legless; fruitfly; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;
 KW tissue proliferation; tumour; cytoskeletal; cellular disorder; colon;
 KW blood disorder; cancer; breast; head and neck cancer; brain; thyroid;
 KW medulloblastoma; skin cancer; tissue regeneration; tissue repair.
 XX
 OS Drosophila melanogaster.
 XX
 XX US2002086986-A1.
 PN
 XX
 PD 04-JUL-2002.
 XX
 XX 27-JUL-2001; 2001US-0915543.
 PF
 XX 28-JUL-2000; 2000US-221502P.
 PR
 XX (BASL/) BASLER K.
 PA (BRUN/) BRUNNER E.
 PA (FROE/) FROESCH B.
 PA (KRAM/) KRAMPS T.
 PA (PETE/) PETER O.
 XX
 XX Basler K, Brunner E, Froesch B, Kramps T, Peter O;
 PI
 XX WPI; 2002-635689/68.
 DR N-PSDB; AAF88466.
 DR
 XX

PT Novel polypeptide useful in therapeutic method for treating disorders
 PT of cell fate such as cell differentiation or cell proliferation -
 PS Example II; Fig 2; 41pp; English.
 XX
 CC This invention describes a novel polypeptide sharing one or more
 CC homologous amino acid domains with the legless (lgs) protein, a
 CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway
 CC involved in the formation and maintenance of spatial arrangements
 CC and proliferation of tissues during development, and in the formation
 CC and growth of many human tumours. The products of the invention have
 CC cytoskeletal activity and can be used to treat cellular disorders, blood
 CC disorders and cancers caused by over-stimulation of the Wnt pathway,
 CC where the cancerous condition is colon, breast, head and neck, brain,
 CC thyroid, medulloblastoma or skin cancer. The product could also be used
 CC to promote tissue regeneration and repair. This sequence represents the
 CC Drosophila melanogaster (fruitfly) legless (lgs) protein described in
 CC the disclosure of the invention.

XX SQ Sequence 1464 AA;

Query Match 72.1%; Score 98; DB 23; Length 1464;
 Best Local Similarity 57.1%; Pred. No. 2.9e-06;
 Matches 16; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
 Db 318 IFVFSTQLANKGAESVLSGQFOTIIAYH 345

RESULT 6

ID AAU78460
 AC AAU78460; standard; Protein; 1494 AA.

02-JUL-2002 (first entry)

Mouse beta-catenin nuclear localised protein.

Mouse; beta-catenin nuclear localised protein; cancer;
 gene therapy; EST; expressed sequence tag.

Mus musculus.

WO200224738-A1.

28-MAR-2002.

19-SEP-2001; 2001WO-JF08140.

22-SEP-2000; 2000JP-0287876.

(KYOW) KYOWA HAKKO KOGYO KK.

Akiyama T, Adachi S;

WPI; 2002-330014/36.

N-PSDB; ABK47631.

New beta-catenin nuclear localised protein for diagnosis and treatment
 of diseases associated with nuclear localisation of beta-catenin e.g.
 cancer -

Claim 1; Page 81-88; 113pp; Japanese.

The invention relates to a beta-catenin nuclear localised protein
 and DNA encoding the protein. The protein and encoding DNA are
 applicable in diagnosis and treatment of diseases associated with
 nuclear localisation of beta-catenin e.g. cancer, including gene
 therapy. The present sequence represents the amino acid sequence of
 mouse beta-catenin nuclear localised protein.

SQ Sequence 1494 AA;

Query Match 69.1%; Score 94; DB 23; Length 1494;
 Best Local Similarity 60.7%; Pred. No. 1.3e-05;
 Matches 17; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
 Db 238 VYVFTHLANTAATAEAVLQGRAESILAYH 265

RESULT 7

ABP06595

ID ABP06595 standard; Protein; 114 AA.

AC ABP06595;

25-JUN-2002 (first entry)

Human ORFX protein sequence SEQ ID NO:13172.

Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 degenerative disorder; osteoarthritis; neurodegenerative disorder;
 cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 hypertension; hypothyroidism; cholesterol ester storage disease;
 immune deficiency; immune disorder; infectious disease;
 autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 myasthenia gravis.

Homo sapiens.

WO200192523-A2.

06-DEC-2001.

29-MAY-2001; 2001WO-US10836.

30-MAY-2000; 2000US-206132P.

29-AUG-2000; 2000US-228716P.

(CURA-) CURAGEN CORP.

Shinkets RA, Leach MD;

WPI; 2002-106308/14.

N-PSDB; ABN22347.

Novel human polypeptides and polynucleotides useful for diagnosing,
 preventing and treating cardiovascular disease, neurodegenerative,
 hyperproliferative disorders and autoimmune disorders -

Disclosure; SEQ ID 13172; 1037pp; English.

The present invention describes substantially purified human proteins
 (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 in the specification). ABN15762 to ABN27252 encode the human ORFX
 proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 treating or preventing a pathology associated with an ORFX-associated
 disorder in humans, and in the manufacture of a medicament for treating a
 syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 sequences can be used in gene therapy. ORFX sequences can be used in the
 treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 osteoarthritis, neurodegenerative disorders, disorders related to organ
 transplantation, cardiovascular diseases, diabetes mellitus, systemic
 lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 storage disease, various immune deficiencies and disorders, infectious
 diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 disease and autoimmune inflammatory eye disease. ORFX proteins are also
 useful for treating burns, incisions, ulcers, for treating osteoporosis,
 bone degenerative disorders, or periodontal disease, and for gut

CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 114 AA;

Query Match 66.9%; Score 91; DB 23; Length 114;
 Best Local Similarity 57.1%; Pred. No. 1.7e-06;
 Matches 16; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

OY 1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28
 ||||| :|| ||||| :|| :|| :||
 Db 78 VYVFTTHLANTAEEAVLQGRADSILAYH 105

RESULT 8

AAU78461
 ID AAU78461 standard; Protein; 320 AA.

XX
 AC AAU78461;

XX
 DT 02-JUL-2002 (first entry)

XX
 DE Mouse beta-catenin nuclear localised protein #2.

XX
 KW Mouse; beta-catenin nuclear localised protein; cancer;
 KW gene therapy; EST; expressed sequence tag.

XX
 OS Mus musculus.

XX
 PN WO200224738-A1.

XX
 PD 28-MAR-2002.

XX
 PF 19-SEP-2001; 2001WO-JP08140.

XX
 PR 22-SEP-2000; 2000JP-0287876.

XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.

XX
 PI Akiyama T, Adachi S;

XX
 DR WPI; 2002-330014/36.

XX
 DR N-PSDB; ABK47632.

XX
 PT New beta-catenin nuclear localised protein for diagnosis and treatment
 PT of diseases associated with nuclear localisation of beta-catenin e.g.
 PT cancer

XX
 PS Claim 2; Page 91-92; 113pp; Japanese.

XX
 CC The invention relates to a beta-catenin nuclear localised protein
 CC and DNA encoding the protein. The protein and encoding DNA are
 CC applicable in diagnosis and treatment of diseases associated with
 CC nuclear localisation of beta-catenin e.g. cancer, including gene
 CC therapy. The present sequence represents the amino acid sequence of
 CC mouse beta-catenin nuclear localised protein #2.

XX
 SQ Sequence 320 AA;

Query Match 49.3%; Score 67; DB 23; Length 320;
 Best Local Similarity 57.1%; Pred. No. 0.035;
 Matches 12; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

OY 8 MANKAAEAVLKQGVETIVSFH 28

Db 1 LANTAAEAVLQGRAESILAYH 21

RESULT 9

AAO07544
 ID AAO07544 standard; Protein; 113 AA.

XX
 AC AAO07544;

XX
 DT 06-NOV-2001 (first entry)

XX
 DE Human polypeptide SEQ ID NO 21436.

XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.

XX
 OS Homo sapiens.

XX
 PN WO200164835-A2.

XX
 PD 07-SEP-2001.

XX
 PF 26-FEB-2001; 2001WO-US04927.

XX
 PR 28-FEB-2000; 2000US-0515126.

XX
 PR 18-MAY-2000; 2000US-0577409.

XX
 PA (HYSE-) HYSEQ INC.

XX
 PI Tang YT, Liu C, Drmanac RT;

XX
 DR WPI; 2001-514838/56.

XX
 DR N-PSDB; AAI87475.

XX
 PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders

XX
 PS Claim 20; SEQ ID NO 21436; 1399pp + Sequence Listing; English.

XX
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.

XX
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 113 AA;

Query Match 47.8%; Score 65; DB 22; Length 113;
 Best Local Similarity 75.0%; Pred. No. 0.021;

Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 8 MANKAAEAVLKQGVETIVSF 27

Db 81 MCYQAAEAVLKQGVETDASF 100

RESULT 10

AAG33446

ID AAG33446 standard; Protein; 360 AA.

XX
 AC AAG33446;

XX
 DT 18-OCT-2000 (first entry)

XX
 DE Zea mays protein fragment SEQ ID NO: 40525.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW Hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX Zea mays subsp. mays.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.

```

PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0155596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 38.2%; Score 52; DB 21; Length 360;
Best Local Similarity 50.0%; Pred. No. 9.4;
Matches 12; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 4 FSTENANKAAEAVALGQVETIVSF 27
Db |||:|:| ||| |||:|:|
97 FSTRLANLENLVLEGPEETAAAF 120

RESULT 11
AAG33445
ID AAG33445 standard; Protein; 448 AA.
XX AC AAG33445;
XX DT 18-OCT-2000 (first entry)
XX DE Zea mays protein fragment SEQ ID NO: 40524.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence; corn.
XX OS Zea mays subsp. mays.
XX

```

```

PN EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.

```

PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	19-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144532.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145218.
PR	26-JUL-1999;	99US-0145224.
PR	27-JUL-1999;	99US-0145376.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	05-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	06-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	07-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	16-AUG-1999;	99US-0148684.
PR	17-AUG-1999;	99US-0148688.
PR	18-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149176.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-SEP-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.

PR	20-SEP-1999;	99US-0154779;
PR	22-SEP-1999;	99US-0155139;
PR	23-SEP-1999;	99US-0155486;
PR	24-SEP-1999;	99US-0155659;
PR	28-SEP-1999;	99US-0156458;
PR	29-SEP-1999;	99US-0156596;
PR	04-OCT-1999;	99US-0157117;
PR	05-OCT-1999;	99US-0157753;
PR	06-OCT-1999;	99US-0157865;
PR	07-OCT-1999;	99US-0158029;
PR	08-OCT-1999;	99US-0158232;
PR	12-OCT-1999;	99US-0158369;
PR	13-OCT-1999;	99US-0159293;
PR	13-OCT-1999;	99US-0159294;
PR	13-OCT-1999;	99US-0159295;
PR	14-OCT-1999;	99US-0159329;
PR	14-OCT-1999;	99US-0159330;
PR	14-OCT-1999;	99US-0159331;
PR	14-OCT-1999;	99US-0159637;
PR	14-OCT-1999;	99US-0159638;
PR	18-OCT-1999;	99US-0159584;
PR	21-OCT-1999;	99US-0160741;
PR	21-OCT-1999;	99US-0160767;
PR	21-OCT-1999;	99US-0160768;
PR	21-OCT-1999;	99US-0160770;
PR	21-OCT-1999;	99US-0160814;
PR	21-OCT-1999;	99US-0160815;
PR	22-OCT-1999;	99US-0160800;
PR	22-OCT-1999;	99US-0160981;
PR	22-OCT-1999;	99US-0160989;
PR	25-OCT-1999;	99US-0161404;
PR	25-OCT-1999;	99US-0161405;
PR	25-OCT-1999;	99US-0161406;
PR	26-OCT-1999;	99US-0161359;
PR	26-OCT-1999;	99US-0161360;
PR	26-OCT-1999;	99US-0161361;
PR	28-OCT-1999;	99US-0161920;
PR	28-OCT-1999;	99US-0161992;
PR	28-OCT-1999;	99US-0162193;
PR	29-OCT-1999;	99US-0162142;

Query Match 38.2%; Score 52; DB 21; Length 448;

Best Local Similarity 50.0%; Pred. No. 12;
Matches 12; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKGQVETIVSF 27
111 : 11 111 111 : 1

Db 185 FSTRIANNLLENVLKQGPETIAAF 208

RESULT 12

RESUL 12
AAG33444
ID AAG33444 standard: Protein: 509 AA.

XX
AC

DT 18-OCT-2000 (first entry)

DE Zea mays protein fragment SEQ ID NO: 40523.

XX
KW protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.

OS Zea mays subsp. mays.

XX
PN EP1033405-A2

XX
PD 06-SEP-2000.

XX
PF 25-FEB-2000; 2000EP-0301439.

PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127482.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.

05-OCT-1999	9905-0157753
06-OCT-1999	9905-0157865
07-OCT-1999	9905-0158029
08-OCT-1999	9905-0158232
12-OCT-1999	9905-0158369
13-OCT-1999	9905-0159293
13-OCT-1999	9905-0159294
13-OCT-1999	9905-0159295
14-OCT-1999	9905-0159329
14-OCT-1999	9905-0159330
14-OCT-1999	9905-0159331
14-OCT-1999	9905-0159637
14-OCT-1999	9905-0159638
18-OCT-1999	9905-0159584
21-OCT-1999	9905-0160741
21-OCT-1999	9905-0160767
21-OCT-1999	9905-0160768
21-OCT-1999	9905-0160770
21-OCT-1999	9905-0160778
21-OCT-1999	9905-0160814
21-OCT-1999	9905-0160815
22-OCT-1999	9905-0160980
22-OCT-1999	9905-0160981
22-OCT-1999	9905-0160980
25-OCT-1999	9905-0160989
25-OCT-1999	9905-0161404
25-OCT-1999	9905-0161405
25-OCT-1999	9905-0161406
26-OCT-1999	9905-0161359
26-OCT-1999	9905-0161360
26-OCT-1999	9905-0161361
28-OCT-1999	9905-0161920
28-OCT-1999	9905-0161921
28-OCT-1999	9905-0161922
28-OCT-1999	9905-0161923
29-OCT-1999	9905-0162142

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 7953; 2lpp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CCC capable of detecting 1000 or more genes from Drosophila. The invention is
CCC useful in developmental biology and in elucidating cell signalling and
CCC cell-cell interactions in higher eukaryotes for the development of
CCC insecticides, therapeutics and pharmaceutical drugs. The invention
CCC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CCC sequences (ABL01840-ABL16175) and the encoded proteins
CCC (ABB57737-ABB72072).
CCC The sequence data for this patent did not form part of the printed
CCC specification, but was obtained in electronic format directly from WIPO
CCC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 1049 AA;

Query Match 37.5%; Score 51; DB 22; Length 1049;
Best Local Similarity 40.7%; Pred. NO. 49;
Matches 11; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 2 YVFSTENAKAAEAVLKGVETIVSFH 28
|| :||| ||| ::|||)
DB 471 YVSCDMAASATNVRSGELKIPIEH 497

RESULT 14
ABB89793
ID ABB89793 standard; Protein; 130 AA.
XX
XX AC ABB89793;
XX
DT 24-MAY-2002 (first entry)
XX
XX Human polypeptide SEQ ID NO 2169.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnary; anticoagulant; antibacterial; antifungal; antiparasitic;
KW cardiarg; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX
OS Homo sapiens.
XX
PN WO200190304-A2.
XX
PD 29-NOV-2001.
XX
PF 18-MAY-2001; 2001WO-US16450.
XX
PR 19-MAY-2000; 2000US-205515P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
XX
DR WPI: 2002-122018/16.
DR N-PSSB; ABL90202.
XX
PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders -
XX
PS Claim 11; SEQ ID NO 2169; 2081pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 5.66537 Seconds
(without alignments)
209.113 Million cell updates/sec

Title: US-09-915-543-15_COPY_177_204

Perfect score: 136

Sequence: 1 VYVFSTEMANKAAEAVLKQVETIVSFH 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	34.6	2639	3	US-09-080-983-3
2	44	32.4	718	3	US-09-090-808-2
3	44	32.4	718	3	US-09-447-453-2
4	43	31.6	145	4	US-09-091-725-37
5	43	31.6	438	3	US-08-486-099-93
6	43	31.6	438	3	US-08-360-107A-103
7	43	31.6	438	3	US-08-484-223B-93
8	43	31.6	438	3	US-08-919-597-93
9	43	31.6	438	3	US-08-475-668A-93
10	43	31.6	438	3	US-08-485-551A-93
11	43	31.6	438	3	US-08-471-913A-93
12	43	31.6	438	3	US-08-485-264A-93
13	43	31.6	438	4	US-08-474-349A-93
14	43	31.6	438	4	US-08-255-208A-29
15	43	31.6	438	4	US-08-470-896-93
16	43	31.6	438	4	US-08-485-546A-93
17	43	31.6	662	1	US-08-224-657-88
18	43	31.6	662	4	US-09-354-138-88
19	42	30.9	154	4	US-09-732-210-103
20	42	30.9	154	4	US-09-732-210-210
21	42	30.9	188	4	US-09-252-991A-29737
22	42	30.9	300	4	US-09-634-238-219
23	42	30.9	358	4	US-09-934-901-8
24	42	30.9	541	4	US-09-627-376-11
25	42	30.9	1820	3	US-07-998-289B-8
26	41.5	30.5	111	4	US-09-775-932-20
27	41.5	30.5	678	1	US-08-844-085-2

28 41.5 30.5 745 1 US-08-453-472-5 Sequence 5, Appli
29 41.5 30.5 745 1 US-08-038-948-9 Sequence 9, Appli
30 41.5 30.5 745 1 US-08-453-952-5 Sequence 5, Appli
31 41.5 30.5 745 2 US-08-484-993B-43 Sequence 5, Appli
32 41.5 30.5 745 2 US-08-862-903-5 Sequence 43, Appli
33 41.5 30.5 745 2 US-08-484-158B-43 Sequence 43, Appli
34 41.5 30.5 745 2 US-08-484-596A-43 Sequence 43, Appli
35 41.5 30.5 745 2 US-08-480-150A-43 Sequence 43, Appli
36 41.5 30.5 745 3 US-08-458-731-43 Sequence 43, Appli
37 41.5 30.5 745 3 US-08-149-223A-43 Sequence 43, Appli
38 41 30.1 111 4 US-09-107-532A-5449 Sequence 5449, Ap
39 41 30.1 162 4 US-09-134-001C-3670 Sequence 1, Appli
40 41 30.1 211 3 US-09-075-194-1 Sequence 6, Appli
41 41 30.1 256 2 US-08-578-516-6 Sequence 6500, Ap
42 41 30.1 378 4 US-09-107-532A-6500 Sequence 6689, Ap
43 41 30.1 569 4 US-09-107-532A-6689 Sequence 5806, Ap
44 41 30.1 676 4 US-09-107-532A-5806 Sequence 1, Appli
45 41 30.1 689 4 US-09-499-964-1

ALIGNMENTS

RESULT 1
US-09-080-983-3
; Sequence 3, Application US/09080983
; Patent No. 6197948
; GENERAL INFORMATION:
; APPLICANT: Zhu, Hai-Ying
; APPLICANT: Ling, Kai-Shu
; APPLICANT: Gonsalves, Dennis
; TITLE OF INVENTION: GRAPEVINE LEAFROLL VIRUS TYPE 2 PROTEINS
; TITLE OF INVENTION: AND THEIR USES
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080.983
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/047,194
; FILING DATE: 20-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1631
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2639 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-080-983-3

Query Match 34.6%; Score 47; DB 3; Length 2639;
Best Local Similarity 33.3%; Pred. No. 1.3e+02;
Matches 9; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

```

, COUNTRY: United States of America
, ZIP: 20006-1888
,
, COMPUTER READABLE FORM:
,
, MEDIUM TYPE: Floppy disk
,
, COMPUTER: IBM PC compatible
,
, OPERATING SYSTEM: PC-DOS/MS-DOS
,
, SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
,
, CURRENT APPLICATION DATA:
,
, APPLICATION NUMBER: US/09/091,725
,
, FILING DATE: 23-DEC-1996
,
, CLASSIFICATION: 435
,
, PRIOR APPLICATION DATA:
,
, APPLICATION NUMBER: EP 95203620.0
,
, FILING DATE: 22-DEC-1995
,
, APPLICATION NUMBER: EP 96200943.7
,
, FILING DATE: 11-APR-1996
,
, ATTORNEY/AGENT INFORMATION:

```

```
; NAME: E. Victor Donahue
; REGISTRATION NUMBER: 35,492
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 145 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-091-725-37

Query Match 31.6%; Score 43; DB 4; Length 145;
Best Local Similarity 43.8%; Pred. No. 16;
Matches 14; Conservative 4; Mismatches 6; Indels 8; Gaps 2;

QY 5 STEMANKAA-EAVLKG-----QVETIVSPH 28
   1 : : ||| : ||| :
   15 SADVKAAKAALKAGTQSTSTRKVRTSVSFH 46
Db

RESULT 5
US-08-486-099-93
; Sequence 93, Application US/08486099
; Patent No. 6013263
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HEPATITIS
; TITLE OF INVENTION: B VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,099
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-031
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-486-099-93

Query Match 31.6%; Score 43; DB 3; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKGQVETIVS 26
Db
```

```
Db 41 SLEQSNKAIEIREATQETVIA 62
   1 : : ||| : : ||| :
   41 SLEQSNKAIEIREATQETVIA 62

RESULT 6
US-08-360-107A-103
; Sequence 103, Application US/08360107A
; Patent No. 6017536
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 149
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/360,107A
; FILING DATE: 20-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 103:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-360-107A-103

Query Match 31.6%; Score 43; DB 3; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKGQVETIVS 26
   1 : : ||| : : ||| :
   41 SLEQSNKAIEIREATQETVIA 62
Db

RESULT 7
US-08-484-223B-93
; Sequence 93, Application US/08484223B
; Patent No. 6020459
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
```

```

; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 245
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,223B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-484-223B-93

Query Match 31.6%; Score 43; DB 3; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKQGVETIVS 26
| | : | | | | : | | : |
Db 41 SLEQSNKAIEIREATQETVIA 62

RESULT 8
US-08-919-597-93
; Sequence 93, Application US/08919597
; Patent No. 6054265
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/470,896
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-919-597-93

Query Match 31.6%; Score 43; DB 3; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKQGVETIVS 26
| | : | | | | : | | : |
Db 41 SLEQSNKAIEIREATQETVIA 62

RESULT 9
US-08-475-668A-93
; Sequence 93, Application US/08475668A
; Patent No. 6060065
; GENERAL INFORMATION:
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 211
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,668A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-475-668A-93

Query Match      31.6%; Score 43; DB 3; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY      5 STEMANKAAEAVLKQGVETIVS 26
Db      41 SLEQSNKAIEIREATQETVIA 62

RESULT 10
US-08-485-551A-93
; Sequence 93, Application US/08485551A
; Patent No. 6068973
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
; NUMBER OF SEQUENCES: 211
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-485-551A-93

Query Match      31.6%; Score 43; DB 3; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY      5 STEMANKAAEAVLKQGVETIVS 26
Db      41 SLEQSNKAIEIREATQETVIA 62

RESULT 12
US-08-485-264A-93
; Sequence 93, Application US/08485264A
; Patent No. 6228983
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING

```

```

RESULT 11
US-08-471-913A-93
; Sequence 93, Application US/08471913A
; Patent No. 6093794
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS
; NUMBER OF SEQUENCES: 214
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-471-913A-93

Query Match      31.6%; Score 43; DB 3; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY      5 STEMANKAAEAVLKQGVETIVS 26
Db      41 SLEQSNKAIEIREATQETVIA 62

RESULT 12
US-08-485-264A-93
; Sequence 93, Application US/08485264A
; Patent No. 6228983
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING

```


RESULT 13
US-08-474-349A-93
; Sequence 93, Application US/08474349A
; Patent No. 633395
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING HUMAN PARAINFLUENZA
; NUMBER OF SEQUENCES: 517
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

INVENTOR: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
APPLICANT: Barney, Shawn O.
APPLICANT: Lambert, Dennis M.
APPLICANT: Petteway Jr., Stephen R.
TITLE OF INVENTION: SYNTHETIC PEPTIDE INHIBITORS OF HIV
TRANSMISSION
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York

ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/255,208A
FILING DATE: 07-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 438 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein

US-08-255-208A-29

Query Match 31.6%; Score 43; DB 4; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKQGVETIVS 26
| | : || | | : : || : :
DB 41 SLEQSNKAIEIREATQETVIA 62

RESULT 15

US-08-470-896-93
; Sequence 93, Application US/08470896
; Patent No. 6479055
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,896
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-470-896-93

Query Match 31.6%; Score 43; DB 4; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKQGVETIVS 26
| | : || | | : : || : :
DB 41 SLEQSNKAIEIREATQETVIA 62

Search completed: October 15, 2003, 10:31:58
Job time : 6.66537 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:32:06 ; Search time 11.8755 Seconds
(without alignments)
379.908 Million cell updates/sec

Title: US-09-915-543-15_COPY_177_204

Perfect score: 136

Sequence: 1 VYVSTEMANKAAEAVLKQGVETIVSFH 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 600653 seqs, 161128416 residues

Total number of hits satisfying chosen parameters: 600653

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136	100.0	28	15	US-10-322-579-3
2	136	100.0	1426	15	US-10-322-579-15
3	98	72.1	28	15	US-10-322-579-2
4	49	36.0	984	15	US-10-128-714-3547
5	49	36.0	1058	15	US-10-128-714-8547
6	47	34.6	732	14	US-10-008-355-6
7	46	33.8	471	12	US-10-167-547C-8
8	45	33.1	507	12	US-10-167-547C-10
9	45	33.1	747	15	US-10-156-761-10721
10	44.5	32.7	170	12	US-10-342-224-92
11	44	32.4	171	15	US-10-156-761-9002
12	44	32.4	484	9	US-09-815-242-10457
13	44	32.4	718	14	US-10-117-846-2
14	43	31.6	80	11	US-09-764-891-5144
15	43	31.6	315	12	US-10-306-762-212

16	43	31.6	623	12	US-10-032-585-7819	Sequence 7819, Ap
17	43	31.6	662	11	US-09-951-061A-141	Sequence 141, App
18	42	30.9	248	12	US-10-097-111-304	Sequence 304, App
19	42	30.9	277	15	US-10-156-761-8036	Sequence 8036, Ap
20	42	30.9	326	15	US-10-156-761-9336	Sequence 9336, Ap
21	42	30.9	358	10	US-09-934-901-8	Sequence 8, Appli
22	42	30.9	358	10	US-09-934-868-18	Sequence 18, Appl
23	42	30.9	358	12	US-10-320-874-8	Sequence 8, Appli
24	42	30.9	358	16	US-10-320-924-8	Sequence 8, Appli
25	42	30.9	380	12	US-10-183-708-42	Sequence 42, Appl
26	42	30.9	380	12	US-09-932-227-42	Sequence 42, Appl
27	42	30.9	399	12	US-10-293-971-9	Sequence 9, Appli
28	42	30.9	541	14	US-10-047-676A-11	Sequence 11, Appl
29	41.5	30.5	111	10	US-09-775-932-20	Sequence 20, Appl
30	41.5	30.5	469	8	US-08-761-986A-5245	Sequence 5245, Ap
31	41	30.1	227	9	US-09-815-242-11197	Sequence 11197, A
32	41	30.1	279	9	US-09-861-451A-28	Sequence 28, Appl
33	41	30.1	370	10	US-09-712-363-187	Sequence 187, App
34	41	30.1	370	12	US-10-361-460-10	Sequence 10, Appl
35	41	30.1	451	11	US-09-910-186A-12	Sequence 12, Appl
36	41	30.1	456	9	US-09-815-242-4931	Sequence 4931, Ap
37	41	30.1	456	9	US-09-815-242-10636	Sequence 10636, A
38	41	30.1	516	10	US-09-738-626-5086	Sequence 5086, Ap
39	41	30.1	568	15	US-10-156-761-11478	Sequence 11478, A
40	41	30.1	638	15	US-10-234-432-59	Sequence 59, Appl
41	41	30.1	689	14	US-10-052-664-1	Sequence 1, Appli
42	41	30.1	689	15	US-10-097-340-221	Sequence 221, App
43	41	30.1	693	9	US-09-741-669-393	Sequence 393, App
44	41	30.1	727	15	US-10-234-432-30	Sequence 30, Appl
45	41	30.1	727	15	US-10-234-432-57	Sequence 57, Appl

ALIGNMENTS

RESULT 1
US-10-322-579-3
; Sequence 3, Application US/10322579
; Publication No. US20030114413A1
; GENERAL INFORMATION:
; APPLICANT: BASLER, Konrad
; APPLICANT: BRUNNER, Erich
; APPLICANT: FROESCH, Barbara
; APPLICANT: KRAMPS, Thomas
; APPLICANT: PETER, Oliver
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATH
; FILE REFERENCE: 050361
; CURRENT APPLICATION NUMBER: US/10/322,579
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US/09/915,543
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 60/221,502
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Human lgs/dcl9
US-10-322-579-3

Query Match 100.0%; Score 136; DB 15; Length 28;
Best Local Similarity 100.0%; Pred. No. 8.3e-15;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYVSTEMANKAAEAVLKQGVETIVSFH 28

DB 1 VYVSTEMANKAAEAVLKQGVETIVSFH 28

RESULT 2

US-10-322-579-15

```
; Sequence 15, Application US/10322579
; Publication No. US20030114413A1
; GENERAL INFORMATION:
; APPLICANT: BASLER, Konrad
; APPLICANT: BRUNNER, Erich
; APPLICANT: FROESCH, Barbara
; APPLICANT: KRAMPS, Thomas
; APPLICANT: PETER, Oliver
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY
; FILE REFERENCE: Q60361
; CURRENT APPLICATION NUMBER: US/10/322,579
; PRIOR FILING DATE: 2002-12-19
; PRIOR FILING DATE: 2001-07-27
; PRIOR FILING DATE: 2001-07-27
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 1426
; TYPE: PRN
; ORGANISM: Human lgs/bc19
US-10-322-579-15

Query Match      100.0%; Score 136; DB 15; Length 1426;
Best Local Similarity 100.0%; Pred. No. 9.4e-13;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKGOVETIVSFH 28
Db 177 VYVFSTEMANKAAEAVLKGOVETIVSFH 204

RESULT 3
US-10-322-579-2
; Sequence 2, Application US/10322579
; Publication No. US20030114413A1
; GENERAL INFORMATION:
; APPLICANT: BASLER, Konrad
; APPLICANT: BRUNNER, Erich
; APPLICANT: FROESCH, Barbara
; APPLICANT: KRAMPS, Thomas
; APPLICANT: PETER, Oliver
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY
; FILE REFERENCE: Q60361
; CURRENT APPLICATION NUMBER: US/10/322,579
; PRIOR FILING DATE: 2002-12-19
; PRIOR FILING DATE: 2001-07-27
; PRIOR FILING DATE: 2001-07-27
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 28
; TYPE: PRN
; ORGANISM: Drosophila lgs
US-10-322-579-2

Query Match      72.1%; Score 98; DB 15; Length 28;
Best Local Similarity 57.1%; Pred. No. 9.6e-09;
Matches 16; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKGOVETIVSFH 28
Db 1 IFVFTQLANKGAESVLSQFQTIIAYH 28

RESULT 4
US-10-128-714-3547
; Sequence 3547, Application US/10128714
```

```
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; PRIOR FILING DATE: 2002-04-23
; PRIOR FILING DATE: 2001-04-23
; PRIOR FILING DATE: 2001-04-23
; PRIOR FILING DATE: 2001-04-27
; PRIOR FILING DATE: 2001-06-05
; PRIOR FILING DATE: 2001-07-09
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3547
; LENGTH: 984
; TYPE: PRN
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3547

Query Match      36.0%; Score 49; DB 15; Length 984;
Best Local Similarity 52.8%; Pred. No. 46;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 VFSTEMANKAAEAVLKGV 21
Db 500 VYRANWANKSAAAVLKSKL 518

RESULT 5
US-10-128-714-8547
; Sequence 8547, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; PRIOR FILING DATE: 2002-04-23
; PRIOR FILING DATE: 2001-04-23
; PRIOR FILING DATE: 2001-04-27
; PRIOR FILING DATE: 2001-04-27
; PRIOR FILING DATE: 2001-06-05
; PRIOR FILING DATE: 2001-07-09
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8547
; LENGTH: 1058
; TYPE: PRN
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8547
```

```
Query Match          36.0%; Score 49; DB 15; Length 1058;
Best Local Similarity 52.6%; Pred. No. 50;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 VFSTEMANKAAEAVLKGV 21
   I: ||||| ||||| :
Db 500 VYRANMANKSAAAVLKSKL 518

RESULT 6
US-10-008-355-6
; Sequence 6, Application US/10008355
; Publication No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Bandula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235,00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 732
; TYPE: PRT
; ORGANISM: Shewanella putrefaciens
US-10-008-355-6

Query Match          34.6%; Score 47; DB 14; Length 732;
Best Local Similarity 39.1%; Pred. No. 67;
Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKGV 23
   I: | | | | | : | : |
Db 517 VALYDTNNAQEAKELAGLST 539

RESULT 7
US-10-167-547C-8
; Sequence 8, Application US/10167547C
; Publication No. US20030170653A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Company
; APPLICANT: Damude, Howard G.
; TITLE OF INVENTION: A Biological Method for the Production of Alpha-Methylene-Gamma
; FILE REFERENCE: CL1804 US NA
; CURRENT APPLICATION NUMBER: US/10/167,547C
; CURRENT FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: 60/297198
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Microsoft Office 07
; SEQ ID NO 8
; LENGTH: 471
; TYPE: PRT
; ORGANISM: tulip pistil
US-10-167-547C-8

Query Match          33.8%; Score 46; DB 12; Length 471;
Best Local Similarity 41.7%; Pred. No. 57;
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKGV 27
   ||| : || : | ||| :
Db 204 FSTRILANLEKILTEGETIAAF 227

RESULT 8
US-10-167-547C-10
; Sequence 10, Application US/10167547C
; Publication No. US20030170653A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Company
; APPLICANT: Damude, Howard G.
; TITLE OF INVENTION: A Biological Method for the Production of Alpha-Methylene-Gam
; FILE REFERENCE: CL1804 US NA
; CURRENT APPLICATION NUMBER: US/10/167,547C
; CURRENT FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: 60/297198
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Microsoft Office 07
; SEQ ID NO 10
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Alstroemeria
US-10-167-547C-10

Query Match          33.1%; Score 45; DB 12; Length 507;
Best Local Similarity 37.5%; Pred. No. 90;
Matches 9; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKGV 27
   ||| : || : | ||| :
Db 246 FSTRILANLEKILTEGETIAAF 269

RESULT 9
US-10-156-761-10721
; Sequence 10721, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-282
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10721
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10721

Query Match          33.1%; Score 45; DB 15; Length 747;
Best Local Similarity 39.1%; Pred. No. 1,4e+02;
Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 3 VFSTEMANKAAEAVLKGV 25
   ||| : || : | ||| :
Db 250 VFATSSLNRRARQALDRGETERAV 272

RESULT 10
US-10-342-224-92
; Sequence 92, Application US/10342224
; Publication No. US20030162294A1
; GENERAL INFORMATION:
; APPLICANT: Nathalie Verbruggen
; TITLE OF INVENTION: Genes Involved in Tolerance to Environmental Stress
```

```

; FILE REFERENCE: CNN-012US
; CURRENT APPLICATION NUMBER: US/10/342,224
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: US/09/762,154
; PRIOR FILING DATE: 2002-02-02
; PRIOR APPLICATION NUMBER: EP 98202634.6
; PRIOR FILING DATE: 1998-08-04
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 92
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-342-224-92

Query Match          32.7%; Score 44.5; DB 12; Length 170;
Best Local Similarity 52.0%; Pred. No. 29;
Matches 13; Conservative 0; Mismatches 7; Indels 5; Gaps 1;

QY 6 TEMANKAAEAVL-----KGQVETIV 25
Db 85 TEYVRKTVVVLDTLEKKQVETIV 109

RESULT 11
US-10-156-761-9002
; Sequence 9002, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9002
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9002

Query Match          32.4%; Score 44; DB 15; Length 171;
Best Local Similarity 35.7%; Pred. No. 35;
Matches 10; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
Db 29 IFCIRASMAAKAEQVLVDQWRGILAVH 56

RESULT 12
US-09-815-242-10457
; Sequence 10457, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

```

```

; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA_011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10457
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10457

Query Match          32.4%; Score 44; DB 9; Length 484;
Best Local Similarity 30.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 8 MANKAAEAVLKGQVETIVSF 27
Db 52 LLRKSVSVIINGTIKTIIGF 71

RESULT 13
US-10-117-846-2
; Sequence 2, Application US/10117846
; Publication No. US20020168673A1
; GENERAL INFORMATION:
; APPLICANT: Fuller, Margaret T
; APPLICANT: Hales, Karen G.
; APPLICANT: Santel, Ansgar H.
; TITLE OF INVENTION: Mitofusins, Fzo Homologs and Functional
; FILE REFERENCE: STAN-063CIP3
; CURRENT APPLICATION NUMBER: US/10/117,846
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: 09/413,285
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: PCT/US00/27871
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 718
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-117-846-2

Query Match          32.4%; Score 44; DB 14; Length 718;
Best Local Similarity 34.6%; Pred. No. 2e+02;
Matches 9; Conservative 5; Mismatches 10; Indels 2; Gaps 1;

QY 2 YVFS--TEMANKAAEAVLKGQVETIV 25
Db 403 YCFEELTEMTQVRGRCVLSQIKTLI 428

RESULT 14
US-09-764-891-5144
; Sequence 5144, Application US/09764891

```


Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	60	44.1	205	2	I40812		porphobilinogen sy
2	55	40.4	634	2	T27465		hypothetical prote
3	51	37.5	243	2	G97070		zn-dependent hydro
4	49	36.0	210	2	D86398		protein F17L21.2 [
5	49	36.0	458	2	F71315		probable response
6	49	36.0	586	2	D82484		Sgat protein VCA02
7	48.5	35.7	311	2	C81380		probable D-2-hydro
8	48	35.3	330	2	E84074		dihydroxyacetone k
9	47	34.6	319	2	T01822		hypothetical prote
10	47	34.6	352	2	F90179		hypothetical prote
11	47	34.6	662	2	AD0623		prolidase (Xaa-Pro
12	46.5	34.2	586	2	T23695		probable bacteriop
13	46	33.8	330	2	S08500		hypothetical prote
14	46	33.8	363	2	C65962		QUNc protein - Eme
15	46	33.8	461	2	E95887		branched-chain fat
16	46	33.8	1289	2	E95887		probable aminotran
17	46	33.8	2123	2	F72308		hypothetical prote
18	45.5	33.5	256	2	S55089		probable acetyl-Co
19	45.5	33.5	268	2	SA0363		class II histocomp
20	45.5	33.5	504	2	E92961		glutamate CoA-tr
21	45.5	33.5	505	2	S54744		cellulase (EC 3.2.
22	45	33.1	131	2	S39962		endoglucanase - Er
23	45	33.1	451	2	H72478		hypothetical prote
24	45	33.1	555	2	B96495		hypothetical prote
25	44.5	32.7	48	2	H96762		hypothetical prote
26	44.5	32.7	50	2	D90907		hypothetical prote
27	44	32.4	126	2	F85710		unknown protein en
28	44	32.4	265	2	T43131		hypothetical prote
29	44	32.4	318	2	T14645		hypothetical prote
					C64444		conserved hypothet

QY 2 YVFSTEMANKAAEAVLKGQVETIVSFH 28
 ||| ||| || | : : | ||
 Db 461 YVKCAHMAEKAAVAANGDLQIIPFHH 487

RESULT 3
 G97070
 Zn-dependent hydrolases, glyoxylase family [imported] - Clostridium acetobutylicum
 C:Species: Clostridium acetobutylicum
 C:date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C:Accession: G97070
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
 J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A:title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl
 A:Reference number: A96900; PMID:21359325; PMID:21359325
 A:Accession: G97070
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-243 <KUR>
 A:Cross-references: GB:AE001437; PIDN:AAK79354.1; PID:g15024323; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC1386

Query Match 37.5%; Score 51; DB 2; Length 243;
 Best Local Similarity 25.0%; Pred. No. 5.9;
 Matches 7; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

QY 1 YVFSTEMANKAAEAVLKGQVETIVSFH 28
 ::||::||::||::||::||::||::||::||
 Db 199 LFDFSNLSKKSLKLYDIETVICYH 226

RESULT 4
 D86398
 protein F17L21.2 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: D86398
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Luo, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; PMID:21016719; PMID:11130712
 A:Accession: D86398
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-210 <STO>
 A:Cross-references: GB:AE005172; NID:g9802520; PIDN:AAF99722.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F17L21.2
 A:Map position: 1

Query Match 36.0%; Score 49; DB 2; Length 210;
 Best Local Similarity 41.7%; Pred. No. 10;
 Matches 10; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKGQVETIVSF 27
 :|||::||::||::||
 Db 181 WFSRSTNKADRLAKGELENNVTFF 204

RESULT 5
 F71315
 probable response regulatory protein (atoC) - syphilis spirochete

A:Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyp

A:Reference number: AB1250; MUID:20150912; PMID:10688204

A:Accession: C81380

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-311 <PAR>

A:Cross-references: GB:AL111168; GB:AL1139075; GB:AL111168; NID:96967817; PIDN:CAB74209.1; PID:9696784

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

C:Superfamily: phosphoglycerate dehydrogenase

Query Match 35.7%; Score 48.5; DB 2; Length 311;

Best Local Similarity 42.4%; Pred. No. 18;

Matches 14; Conservative 6; Mismatches 8; Indels 5; Gaps 2;

QY 1 VYFSTEMANKAAEAV---LKGQVET--IVSFH 28

Db 171 IYYSTSGANKNADFVHLEKLDLKYCTDIISIH 203

RESULT 8

E84074

A:Title: dihydroxyacetone kinase BH3397 [imported] - *Bacillus halodurans* (strain C-125)

C:Species: *Bacillus halodurans*

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: E84074

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and

A:Reference number: AB3650; MUID:20512582; PMID:11058132

A:Accession: E84074

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-330 <STO>

A:Cross-references: GB:AP001518; GB:BA000004; NID:g10175792; PIDN:BA07116.1; GSPDB:GN00

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH3397

Query Match 35.3%; Score 48; DB 2; Length 330;

Best Local Similarity 42.3%; Pred. No. 23;

Matches 11; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 2 YVFSTEMANKAAEAVLKGQVETIVSF 27

Db 270 YVPMNDVANKLTREGLNIQFKVGSF 295

RESULT 9

T01822

A:Title: hypothetical protein T27D20.16 - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 10-Dec-1999

C:Accession: T01822

R;Edwards, J.; Wollam, C.; Dubbelde, C.

submitted to the EMBL data Library, August 1998

A:Description: The sequence of A. thaliana T27D20.

A:Reference number: Z1441

A:Accession: T01822

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-319 <EDW>

A:Cross-references: EMBL:AF076274; NID:g3293583; PID:g3377852

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 4

A:Introns: 49/3; 151/3; 210/3; 269/2

A:Note: T27D20.16

C:Superfamily: *Arabidopsis* hypothetical protein F7N22.18

Query Match 34.6%; Score 47; DB 2; Length 319;

Best Local Similarity 48.0%; Pred. No. 31;

C;Accession: T29695
R;Du, Z.; Gattung, S.
Submitted to the EMBL Data Library, November 1995
A;Description: The sequence of *C. elegans* cosmid T18H9.
A;Reference number: 220666
A;Accession: T29695
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-586 <DUE>
A;Cross-references: EMBL:U41746; PIDN:AAA83334.1; CESP:T18H9.1
C;Genetics:
A;Introns: 23/1; 105/3; 174/3; 303/3; 403/2; 430/3; 539/2

Query Match 34.2%; Score 46.5; DB 2; Length 586;
Best Local Similarity 40.7%; Pred. No. 70;
Matches 11; Conservative 7; Mismatches 8; Indels 1; Gaps 1;

QY 1 VYVFSTEMANKAAEAVLKQGVETIVSF 27
: | | | | : | | | | :
Db 492 VHKPATKL-QKHSEKIFSTQFETIVSY 517

RESULT 13
S08500
QUTG protein - *Emricella nidulans*
C;Species: *Emricella nidulans*, *Aspergillus nidulans*
C;Date: 21-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 24-Sep-1998
C;Accession: S08500; S11944
R;Hawkins, A.R.; Lamb, H.K.; Smith, M.; Keyte, J.W.; Roberts, C.F.
Mol. Gen. Genet. 214, 224-231, 1988
A;Title: Molecular organisation of the quinic acid utilization (QUT) gene cluster in *Asp*
A;Reference number: S08498; MUID:89181521; PMID:2976880
A;Accession: S08500
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-330 <HAW>
A;Cross-references: EMBL:X13525; NID:g23398; PID:g23399
A;Note: the authors translated the codon GCG for residue 327 as Thr
R;Lamb, H.K.; Hawkins, A.R.; Smith, M.; Harvey, I.J.; Brown, J.; Turner, G.; Roberts, C.
Mol. Gen. Genet. 223, 17-23, 1990
A;Title: Spatial and biological characterisation of the complete quinic acid utilisation
A;Reference number: S11944; MUID:91080861; PMID:21175387
A;Accession: S11944
A;Molecule type: DNA
A;Residues: 1-326, 'T' <LAM>
A;Experimental source: strain 2035
C;Genetics:
A;Introns: 57/1; 104/1; 250/2; 293/2

Query Match 33.8%; Score 46; DB 2; Length 330;
Best Local Similarity 38.9%; Pred. No. 45;
Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLK 18
: | | | | : | | :
Db 13 IYAFATDLARKAGQLLE 30

RESULT 14
C69962
branched-chain fatty-acid kinase homolog yqiu - *Bacillus subtilis*
C;Species: *Bacillus subtilis*
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: C69962
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gall
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scan
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, R.; Sekiguchi, J.; Sekowska, A.; S
keucha, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiy
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: C69962
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-363 <KUN>
A;Cross-references: GB:Z99116; GB:AL009126; NID:g2634723; PIDN:CAB14338.1; PID:g2634
A;Experimental source: strain 168
C;Genetics:
A;Gene: yqiu
C;Superfamily: acetate kinase

Query Match 33.8%; Score 46; DB 2; Length 363;
Best Local Similarity 71.4%; Pred. No. 50;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 12 AAEAVLKQGVETIV 25
: | | | | : | | :
Db 293 AASAAKGEVAV 306

RESULT 15
E95887
probable aminotransferase protein [imported] - *Sinorhizobium meliloti* (strain 1021) n
C;Species: *Sinorhizobium meliloti*
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C;Accession: E95887
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; He
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A;Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing
A;Reference number: A95842; MUID:21396508; PMID:11481431
A;Accession: E95887
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-461 <KUR>
A;Cross-references: GB:AL591985; PIDN:CAC48765.1; PID:gl5140238; GSPDB:GN00167
A;Experimental source: strain 1021, megaplasmid pSymB
R;Galibert, F.; Finan, T.M.; Long, S.R.; Publer, A.; Abola, P.; Ampe, F.; Barloy-Hub
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lela
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,
A;Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A;Reference number: A96039; MUID:21368234; PMID:11474104
C;Genetics:
A;Contents: annotation
C;Genetics:
A;Gene: SMD20379
A;Genome: plasmid

Query Match 33.8%; Score 46; DB 2; Length 461;
Best Local Similarity 41.7%; Pred. No. 65;
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKQGVETIVSF 27
: | | | | : | | :
Db 199 FAFMANNLEAILLEGGPEVIAAF 222

Search completed: October 15, 2003, 10:33:07
Job time : 9.64591 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 3.37743 Seconds
(without alignments)
389.867 Million cell updates/sec

Title: US-09-915-543-15_COPY_177_204
Perfect score: 136
Sequence: 1 VYVFTEMANKAAEAVLKQGVETIVSFH 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	136	100.0	1426	1 BCL9_HUMAN	O00512 homo sapien
2	98	72.1	1469	1 BCL9_DROME	Q961d9 drosophila
3	60	44.1	205	1 HEM2_CLOIO	Q92995 clostidium
4	51	37.5	1217	1 SVV_FUGRO	P49696 fugu rubrip
5	46	33.8	330	1 QUTG_EMENI	P25416 emericeila
6	46	33.8	363	1 BUK_BACSU	P54532 bacillus su
7	46	33.8	2273	1 HFA1_YEAST	P32874 saccharomyc
8	45.5	33.5	224	1 SODM_CHAFE	O96347 charybdis f
9	45.5	33.5	444	1 GUNW_ERWCA	O59394 erwinia car
10	45.5	33.5	504	1 GUNW_ERWCA	O59395 erwinia car
11	45.5	33.5	505	1 GUNW_ERWCA	Q47096 erwinia car
12	45.5	33.5	745	1 ZP2_MACRA	O77726 macaca radi
13	45	33.1	490	1 PEM3_ARATH	Q9c6b9 arabidopsis
14	44.5	32.7	170	1 AIG2_ARATH	P54121 arabidopsis
15	44	32.4	318	1 YB64_METJA	Q58564 methanococc
16	44	32.4	340	1 MSCP_CAEEL	P40614 caenorhabdi
17	44	32.4	484	1 DNLI_ECOLI	P39301 escherichia
18	44	32.4	602	1 SNLI_AERPE	Q9ydl8 aeropyrum p
19	44	32.4	629	1 SYR_PYRPU	Q8u149 pyrococcus
20	44	32.4	980	1 SVV_SCHPO	O75005 schizosacch
21	44	32.4	4427	1 PKSL_BACSU	O05470 bacillus su
22	43	31.6	119	1 TH12_CABEL	Q17424 caenorhabdi
23	43	31.6	159	1 RIB4_SCHPO	Q9uub1 schizosacch
24	43	31.6	291	1 EX53_MYCFN	P75403 mycoplasma
25	43	31.6	345	1 TRPD_SULTO	Q37127 sulfolobus
26	43	31.6	537	1 AREH_SCHPO	O10269 schizosacch
27	43	31.6	575	1 YDGE_SCHPO	O10499 schizosacch
28	43	31.6	594	1 SVV_RAT	Q04462 rattus norv
29	43	31.6	609	1 GLSL_HUMAN	Q0u032 homo sapien
30	43	31.6	602	1 NFRV_VIBPR	Q00971 vibrio prot
31	43	31.6	611	1 EMPA_VIBAN	P43147 vibrio angu
32	43	31.6	662	1 VGLF_CDVO	P12569 canine dist
33	43	31.6	895	1 SECA_CYACA	O19911 cyanidium c

RESULT 1

ID	BCL9_HUMAN	STANDARD;	PRT;	1426 AA.
AC	O00512;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	B-cell lymphoma 9 protein (Bcl-9) (Legless homolog).			
GN	BCL9.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Fetal brain;			
RX	MEDLINE=98158621; PubMed=9490669;			
RA	Willis T.G., Zalcborg I.R., Coignet L.J.A., Wlodarska I., Stul M.,			
RA	Jadav D.M., Bastard C., Treleaven J.G., Catovsky D., Silva M.L.M.,			
RA	Dyer M.J.S.;			
RT	"Molecular cloning of translocation t(1;14)(q21;q32) defines a novel			
RT	gene (BCL9) at chromosome 1q21.";			
RL	Blood 91:1873-1881(1998).			
RN	[2]			
RP	FUNCTION.			
RX	MEDLINE=21952490; PubMed=11955446;			
RA	Kramps T., Peter O., Brunner E., Neilen D., Froesch B., Chatterjee S.,			
RA	Murone M., Zuellig S., Basler K.;			
RT	"Wnt/wingless signaling requires BCL9/legless-mediated recruitment of			
RT	pygopus to the nuclear beta-catenin-TCF complex.";			
RL	Cell 109:47-60(2002).			
CC	-!- FUNCTION: Involved in signal transduction through the wnt pathway.			
CC	-!- SUBUNIT: Binds to beta-catenin (CTNBB1), PYGO1 and PYGO2.			
CC	-!- SUBCELLULAR LOCATION: Nuclear (Probable).			
CC	-!- TISSUE SPECIFICITY: Detected at low levels in thymus, prostate,			
CC	testis, ovary and small intestine, and at lower levels in spleen,			
CC	colon and blood.			
CC	-!- DISEASE: Involved in a t(1;14)(q21;q32) chromosomal translocation			
CC	found in a patient with precursor B-cell acute lymphoblastic			
CC	leukemia (ALL). This translocation leaves the coding region			
CC	intact, but may have pathogenic effects due to alterations in the			
CC	expression level of BCL9. Several cases of translocations within			
CC	the 3' untranslated region of BCL9 have been found in B-cell			
CC	malignancies			
CC	-!- CAUTION: It is uncertain whether Met-1 or Met-27 is the initiator.			
CC	-!- CAUTION: Ref.1 sequence differs from that shown due to a			
CC	frameshift in position 1391.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			

ALIGNMENTS

```

DR EMBL; Y13620; CAA73942.1; ALT_FRAME.
DR Genew; HGNC:1008; BCL9.
DR MIM; 602597; -.
DR GO; GO:0007048; P:oncogenesis; TAS.
KW Nuclear protein; Chromosomal translocation; Proto-oncogene;
KW Wnt signaling pathway.
FT DOMAIN 231 1378 PRO-RICH.
FT DOMAIN 347 377 CTNNB1-BINDING.
FT FT DOMAIN 331 335 POLY-PRO 1.
FT FT DOMAIN 514 517 POLY-PRO 2.
FT FT DOMAIN 900 903 POLY-ALA.
FT FT DOMAIN 970 973 POLY-PRO 3.
SQ SEQUENCE 1426 AA; 149314 MW; A240A487716B7F1B CRO64;

Query Match 100.0%; Score 136; DB 1; Length 1426;
Best Local Similarity 100.0%; Pred. No. 3 1e-12;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKQGVETIVSPH 28
Db 177 VYVFSTEMANKAAEAVLKQGVETIVSPH 204

RESULT 2
BCL9_DROME STANDARD; PRT; 1469 AA.
AC Q961D9; Q9V4D2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Bcl-9 homolog (legless protein).
GN BCL9 OR LGS OR CG2041.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.J., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.E., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista A.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,

```

```

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-Berkeley; TISSUE=Embryo;
RP MEDLINE=22426066; PubMed=12537569;
RX Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN [3]
RN SEQUENCE OF 6-1469 FROM N.A., AND MUTAGENESIS OF GLY-514; LEU-534 AND
RP ILE-537.
RX MEDLINE=21952490; PubMed=11955446;
RA Kramps T., Peter O., Brunner E., Neellen D., Froesch B., Chatterjee S.,
RA Murone M., Zuellig S., Basler K.; BCL9/legless-mediated recruitment of
RT "Wnt/wingless signaling requires BCL9/legless-mediated recruitment of
RT pygopus to the nuclear beta-catenin-TCF complex.";
RL Cell 109:47-60(2002).
CC -!- FUNCTION: Involved in signal transduction through the wnt pathway.
CC -!- SUBUNIT: Binds to ARM and PYGO.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically
CC throughout development.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE003844; AAF59345.1; ALT_SEQ.
DR EMBL; AY051651; AAK93075.1; -.
DR EMBL; AF457205; AAL91368.1; -.
DR FlyBase; FBgn0039907; lgs.
DR GO; GO:0005634; C:nucleus; IEP.
DR GO; GO:0030528; F:transcription regulator activity; IPI.
DR GO; GO:0030177; P:positive regulation of wnt receptor signaling; IPI.
DR GO; GO:0007367; P:segment polarity determination; IMP.
KW Nuclear protein; Developmental protein; Segmentation polarity protein;
KW Wnt signaling pathway.
FT DOMAIN 511 555 ARM-BINDING.
FT DOMAIN 1134 1173 ASN-RICH.
FT DOMAIN 1340 1449 GLN-RICH.
FT DOMAIN 1162 1169 POLY-ASN.
FT MUTAGEN 514 514 G->E: IN ALLELE LGS-21L.
FT MUTAGEN 534 534 L->F: IN ALLELE LGS-17E; SEGMENT POLARITY
FT MUTAGEN 537 537 PHENOTYPE.
FT MUTAGEN 537 537 I->K: IN ALLELE LGS-17P.
SQ SEQUENCE 1469 AA; 153759 MW; 5672E01B7200ED08 CRC64;

Query Match 72.1%; Score 98; DB 1; Length 1469;
Best Local Similarity 57.1%; Pred. No. 1.6e-06;
Matches 16; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKQGVETIVSPH 28
Db 323 IFVFSTQANKGAESVLGQFQIIAYH 350

RESULT 3
HEM2_CLOJO STANDARD; PRT; 205 AA.
ID HEM2_CLOJO
AC Q59295;

```

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Delta-aminolevulinic acid dehydratase (EC 4.2.1.24) (Porphobilinogen synthase) (ALAD) (ALADH) (Fragment).
 GN HEMB.
 OS Clostridium josui.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1499;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FERM P-3684;
 RX MEDLINE=95394829; PubMed=7665501;
 RA Fujino E., Fujino T., Karita S., Sakka K., Ohmiya K.;
 RT "Cloning and sequencing of some genes responsible for porphyrin biosynthesis from the anaerobic bacterium Clostridium josui.";
 RL J. Bacteriol. 177:5169-5175(1995).
 CC -!- CATALYTIC ACTIVITY: 2 5-aminolevulinate = porphobilinogen + 2 H(2O).
 CC -!- COFACTOR: Zinc (By similarity).
 CC -!- PATHWAY: Siroheme biosynthesis.
 CC -!- SUBUNIT: Homooctamer (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE ALADH FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D28503; BAA05863.1; -
 DR PIR; I40812; I40812.
 DR HSP; P15002; I94E.
 DR InterPro; IPR001731; Alad_dehydratase.
 DR Pfam; PF00490; ALAD; 1.
 DR PRINTS; PR00144; DALDHYDRATSE.
 DR ProDom; PD002304; Alad_dehydratase; 1.
 DR PROSITE; PS00169; D_ALA_DEHYDRATASE; PARTIAL.
 KW Porphyrin biosynthesis; Lyase; Zinc.
 FT DOMAIN 114 132 ZINC-BINDING (BY SIMILARITY).
 FT NON_TER 205 205
 SQ SEQUENCE 205 AA; 23172 MW; 886F9DAEFDB1144E CRC64;
 Query Match 44.18; Score 60; DB 1; Length 205;
 Best Local Similarity 46.2%; Pred. No. 0.1;
 Matches 12; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
 QY 2 YVFSTEMANKAAEAVLKQVETIVSF 27
 || : | || || || : || : |
 Db 51 YHSPDMVGKAIERAAALKADYKSVLLF 76
 RESULT 4
 SYV_FUGRU
 ID SYV_FUGRU STANDARD; PRT; 1217 AA.
 AC P49696;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Valyl-tRNA synthetase (EC 6.1.1.9) (Valine--tRNA ligase) (ValRS).
 GN VARS1.
 OS Fuqu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Takifugu.
 OX NCBI_TaxID=31033;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97396021; PubMed=9254008;

RA Lim E.H., Corrochano L.M., Elgar G., Brenner S.;
 RT "Genomic structure and sequence analysis of the valyl-tRNA synthetase gene of the Japanese pufferfish, *Fugu rubripes*.";
 RL DNA Seq. 7:141-151(1997).
 CC -!- CATALYTIC ACTIVITY: ATP + L-valine + tRNA(Val) = AMP + diphosphate + L-valyl-tRNA(Val).
 CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC -!- SIMILARITY: THE N-TERMINAL DOMAIN IS SIMILAR TO ELONGATION FACTOR 1-GAMMA.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X91856; CAA62967.1; -
 DR HSP; P96142; IGAX.
 DR InterPro; IPR004046; GST_Cterm.
 DR InterPro; IPR002300; tRNA-synt_1a.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR InterPro; IPR002303; tRNA-synt_val.
 DR Pfam; PF00043; GST_C; 1.
 DR Pfam; PF00133; tRNA-synt_1; 1.
 DR PRINTS; PR00986; TRNASYNTHVAL.
 DR TIGRFAMs; TIGR00422; vals; 1.
 DR PROSITE; PS00178; AA_tRNA_LIGASE_I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
 FT DOMAIN 1 7320 EF-1-GAMMA LIKE.
 FT SITE 293 303 "HIGH" REGION.
 FT SITE 809 813 "KMSKS" REGION.
 FT BINDING 812 812 ATP (BY SIMILARITY).
 SQ SEQUENCE 1217 AA; 138218 MW; 5E08AF24B5C8A7A1 CRC64;
 Query Match 37.5%; Score 51; DB 1; Length 1217;
 Best Local Similarity 37.0%; Pred. No. 14;
 Matches 10; Conservative 8; Mismatches 9; Indels 0; Gaps 0;
 QY 2 YVFSTEMANKAAEAVLKQVETIVSFH 28
 || : | || || || : || : |
 Db 626 YVSCDMGKQAAADVREGRLKIIPDH 652
 RESULT 5
 QUTG_EMENI
 ID QUTG_EMENI STANDARD; PRT; 330 AA.
 AC P25416;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE QUTG protein.
 GN QUTG.
 OS Emericella nidulans (Aspergillus nidulans).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; Emericella.
 OX NCBI_TaxID=162425;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89181521; PubMed=2976880;
 RA Hawkins A.R., Lamb H.K., Smith M., Keyte J.W., Roberts C.F.;
 RT "Molecular organisation of the quinic acid utilization (QUT) gene cluster in *Aspergillus nidulans*.";
 RL Mol. Gen. Genet. 214:224-231(1988).
 RN [2]
 RP SIMILARITY TO INOSITOL MONOPHOSPHATASE.
 RX MEDLINE=91080861; PubMed=2175387;
 RA Lamb H.K., Hawkins A.R., Smith M., Harvey I.J., Brown J., Turner G., Roberts C.F.;
 RT "Spatial and biological characterisation of the complete quinic acid utilisation gene cluster in *Aspergillus nidulans*.";
 RL Mol. Gen. Genet. 223:17-23(1990).

```

RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Pario V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Taconin E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakashi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RL CC -1- CATALYTIC ACTIVITY: ATP + 2-butanote = ADP + butanoyl phosphate.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the acetokinase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D84432; BAA12596.1; -
CC EMBL; Z99116; CAB14338.1; -
CC PIR; C69962; C69962.
CC Subtilist; BG11724; buk.
CC HAMAP; MF_00542; -; 1.
CC InterPro; IPR000890; Acetate_kin.
CC Pfam; PF00871; Acetate_kinase; 1.
CC PRINTS; P00471; ACETATERKINASE.
CC PROSITE; PS01075; ACETATE_KINASE_1; 1.
CC PROSITE; PS01076; ACETATE_KINASE_2; 1.
CC Transferase; Kinase; Complete proteome.
CC KW SEQUENCE 363 AA; 39764 MW; 94ADA51211F98DA0 CRC64;
CC -----
Query Match 33.8%; Score 46; DB 1; Length 363;
Best Local Similarity 71.4%; Pred. No. 23;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 12 ABAVALKGVETIV 25
DB 293 AASAALKEVEALV 306
||| ||||| |||
||| ||||| |||

RESULT 7
HFAL_YEAST
ID HFAL_YEAST STANDARD; PRT; 2273 AA.
AC P32874;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE HFAL protein.
DE GN HFAL OR YMR207C OR YM8261.01C OR YM8325.08C.
OS Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OC NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagers K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII.";

```

```

RL Nature 387:90-93(1997).
RN [2]
RP SEQUENCE OF 125-949 FROM N.A.
RX MEDLINE=94146412; PubMed=7906156;
RA Kearsley S.E.;
RT "Identification of a Saccharomyces cerevisiae gene closely related to
RL FAS3 (acetyl-CoA carboxylase).";
RL DNA Seq. 4:69-70(1993).
CC -!- COFACTOR: BIOTIN (BY SIMILARITY).
CC -!- SIMILARITY: STRONG TO ACETYL-CoA CARBOXYLASE.
CC -!- CAUTION: THE READING FRAME FROM WHICH THIS PROTEIN IN TRANSLATED
CC HAS NO MET INITIATION CODON NEAR TO THE 5'END. IT DOES NOT SEEM TO
CC BE A PSEUDOGENE. THERE ARE NO APPARENT FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z49809; CAA89922.1; -.
DR EMBL; Z48755; CAA88647.1; -.
DR EMBL; Z22558; CAA80280.1; -.
DR PIR; S55089; S55089.
DR HSSP; P24182; LDV1.
DR SGD; S0004820; HPA1.
DR InterPro; IPR001882; Biotin_attach.
DR InterPro; IPR005482; Biotin_carb_C.
DR InterPro; IPR000089; Biotin_lipoyl.
DR InterPro; IPR000022; Carboxyl_trans.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005481; CPase_L_N.
DR Pfam; PF02785; Biotin_carb_C; 1.
DR Pfam; PF00364; Biotin_lipoyl; 1.
DR Pfam; PF00289; Carboxyl_trans; 1.
DR Pfam; PF02089; CPase_L_chain; 1.
DR Pfam; PF02786; CPase_L_D2; 1.
DR PROSITE; PS00188; BIOTIN; 1.
DR PROSITE; PS00866; CPASE_1; 1.
DR PROSITE; PS00867; CPASE_2; 1.
KW Biotin; Ligase; ATP-binding.
FT NP_BIND 332 337 ATP (BY SIMILARITY).
FT ACT_SITE 459 459 BY SIMILARITY.
FT BINDING 804 804 BIOTIN (BY SIMILARITY).
FT CONFLICT 561 561 F -> L (IN REF. 2).
SQ SEQUENCE 2273 AA; 259160 MW; 08727A301549DA92 CRC64;

Query Match 33.8%; Score 46; DB 1; Length 2273;
Best Local Similarity 45.0%; Pred. No. 1.5e+02;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 2 YVFSTEMANKAAEAVLKGV 21
   ||| : : || : |||
Db 659 YVFTEKVRNKYLELLRRGV 678

RESULT 8
SODM_CHAFE
ID SODM_CHAFE STANDARD; PRT; 224 AA.
AC Q96347;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Mn], mitochondrial precursor (EC 1.15.1.1).
OS Charybdis feriatius (Crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubrachyura; Portunoidae; Portunidae; Charybdis.
OX NCBI_TaxID=65693;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Lin C.T., Lai Y.S., Kuo T.J., Chang T.C.;
RT "Molecular cloning, expression, and characterization of a cDNA
RT encoding Mn-superoxide dismutase from crab Charybdis feriatius.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -!- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) -> O(2) + H(2)O(2).
CC -!- COFACTOR: Manganese (By similarity).
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -!- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF019411; AAD01640.1; -.
DR HSSP; P04179; IABM.
DR InterPro; IPR001189; SODismutase.
DR Pfam; PF00081; sodfe; 1.
DR Pfam; PF02777; sodfe_C; 1.
DR ProDom; PD000475; SODismutase; 1.
DR PROSITE; PS00088; SOD_MN; 1.
KW Oxidoreductase; Manganese; Mitochondrion; Transit peptide.
FT TRANSIT 1 20 MITOCHONDRION (BY SIMILARITY).
FT CHAIN 21 224 SUPEROXIDE DISMUTASE [MN].
FT METAL 46 46 MANGANESE (BY SIMILARITY).
FT METAL 94 94 MANGANESE (BY SIMILARITY).
FT METAL 177 177 MANGANESE (BY SIMILARITY).
FT METAL 181 181 MANGANESE (BY SIMILARITY).
SQ SEQUENCE 224 AA; 24527 MW; 107CF19382E9138A CRC64;

Query Match 33.5%; Score 45.5; DB 1; Length 224;
Best Local Similarity 50.0%; Pred. No. 17;
Matches 13; Conservative 3; Mismatches 9; Indels 1; Gaps 1;

QY 2 YVFSTEMA-NKAAEAVLKGVETIVS 26
   || : : ||| ||| ||| |||
Db 54 YVNNLVAEKLAEAKEKGDVSTIIS 79

RESULT 9
GUNN_ERWCA
ID GUNN_ERWCA STANDARD; PRT; 444 AA.
AC Q59394;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase N precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase N)
DE (Cellulase N).
DE CELN.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Atroseptica FCBR C18;
RX MEDLINE=98299944; PubMed=9636315;
RA Olsen O., Thomsen K.K., Weber J., Duus J.O., Svendsen I., Wegener C.,
RA von Wettstein D.;
RT "Transplanting two unique beta-glucanase catalytic activities into
RT one multienzyme, which forms glucose.";
RL Biotechnology 14:71-76(1996).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL

```



```

CC -----
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L39788; AAC37033.1; -.
CC HSP; O85465; IA3H.
CC InterPro; IPR001956; CBD_3.
CC InterPro; IPR001547; Glyco_hydro_5.
CC Pfam; PF00942; CBM3; 1.
CC Pfam; PF00150; cellulase; 1.
CC PROSITE; PS00659; GLYCOSYL-HYDROL_F5; 1.
CC PROSITE; PS001947; CBD_3; 1.
CC CELLULOSE degradation; Hydrolase; Glycosidase; Signal.
CC SIGNAL 1 31 POTENTIAL.
CC CHAIN 32 504
CC FT DOMAIN 32 334
CC FT DOMAIN 335 352
CC FT DOMAIN 353 504
CC FT ACT_SITE 168 168
CC FT ACT_SITE 256 256
CC FT ACT_SITE 256 256
CC SQ SEQUENCE 444 AA; 48300 MW; FAYE4I79004CBB43 CRC64;

Query Match 33.5%; Score 45.5; DB 1; Length 444;
Best Local Similarity 36.7%; Pred. No. 33;
Matches 11; Conservative 5; Mismatches 11; Indels 3; Gaps 1;

QY 2 YVFSTEMANKAAEAVLKQ-----VETIVSFH 28
   I : : : ||| ||| | | : : :
Db 101 YISNP SLANKVKEA VAAQAQSLGVYIIIDWH 130

RESULT 10
GUNW_ERWCA STANDARD; PRT; 504 AA.
ID GUNW_ERWCA
AC Q59395;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase VI precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase VI)
DE (Cellulase VI).
DE CELV1.
GN Erwinia carotovora.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SC3193;
RX MEDLINE=95231512; PubMed=7715600;
RA Mae A., Heikinheimo R., Palva E.T.;
RT "Structure and regulation of the Erwinia carotovora subspecies
RT carotovora SC3193 cellulase gene celv1 and the role of cellulase in
RT phytopathogenicity.";
RL Mol. Gen. Genet. 247:17-26(1995).
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X79241; CAA55823.1; -.
CC PIR; S54744; S54744.

```

```

DR HSP; O85465; IA3H.
DR InterPro; IPR001956; CBD_3.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM3; 1.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL-HYDROL_F5; 1.
DR PROSITE; PS001947; CBD_3; 1.
DR CELLULOSE degradation; Hydrolase; Glycosidase; Signal.
DR SIGNAL 1 31 POTENTIAL.
DR CHAIN 32 504
DR FT DOMAIN 32 334
DR FT DOMAIN 335 352
DR FT DOMAIN 353 504
DR FT ACT_SITE 168 168
DR FT ACT_SITE 256 256
DR FT ACT_SITE 256 256
DR SQ SEQUENCE 504 AA; 54963 MW; OD7ECF74781565FA CRC64;

Query Match 33.5%; Score 45.5; DB 1; Length 504;
Best Local Similarity 36.7%; Pred. No. 38;
Matches 11; Conservative 5; Mismatches 11; Indels 3; Gaps 1;

QY 2 YVFSTEMANKAAEAVLKQ-----VETIVSFH 28
   I : : : ||| ||| | | : : :
Db 101 YIANP SLANKVKEA VAAQAQSLGVYIIIDWH 130

RESULT 11
GUNW_ERWCA STANDARD; PRT; 505 AA.
ID GUNW_ERWCA
AC Q47096;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
DE (Cellulase V).
DE CELV.
GN Erwinia carotovora.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCRI193;
RX MEDLINE=94067016; PubMed=8246888;
RA Cooper V.J.C., Salmond G.P.C.;
RT "Molecular analysis of the major cellulase (CelV) of Erwinia
RT carotovora: evidence for an evolutionary 'mix-and-match' of enzyme
RT domains.";
RL Mol. Gen. Genet. 241:341-350(1993).
CC -!- FUNCTION: Endoglucanase with some exoglucanase activity.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: Has a pH optimum of about 7.0 and a temperature
CC optimum about 42 degrees Celsius.
CC -!- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X76000; CAA53592.1; -.
CC PIR; S39962; S39962.
CC HSP; O85465; IA3H.
CC InterPro; IPR001956; CBD_3.
CC InterPro; IPR001547; Glyco_hydro_5.
CC Pfam; PF00942; CBM3; 1.
CC Pfam; PF00150; cellulase; 1.

```

DR PRODOM; PD001947; CBD_3; 1.
 DR PROSITE; PS00659; GLYCOSYL HYDROL_F5; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 31 POTENTIAL.
 FT CHAIN 32 505 ENDOGLUCANASE V.
 FT DOMAIN 32 334 LINKER.
 FT DOMAIN 335 352 CELLULOSE-BINDING (BY SIMILARITY).
 FT DOMAIN 353 505 CELLULOSE-BINDING (BY SIMILARITY).
 FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
 FT ACT_SITE 505 AA; 54900 MW; DBEA9337BB4D2623 CRC64;
 SQ SEQUENCE 505 AA; 54900 MW; DBEA9337BB4D2623 CRC64;
 Query Match 33.5%; Score 45.5; DB 1; Length 505;
 Best Local Similarity 36.7%; Pred. NO. 38;
 Matches 11; Conservative 5; Mismatches 11; Indels 3; Gaps 1;
 QY 2 YVFSTEMANKAAEAVLKQ---VETIVSFH 28
 I : : : : : | | | | : : : :
 DB 101 YISNPSLANKYKEAVAAQAQSLGVYIIIDWH 130
 RESULT 12
 ZP2_MACRA STANDARD; PRT; 745 AA.
 ID ZP2_MACRA
 AC Q77726;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Zonae pellucida sperm-binding protein 2 precursor (Zonae pellucida glycoprotein ZP2) (Zonae pellucida protein A).
 GN ZP2.
 OS Macaca radiata (Bonnet monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 OX NCBI_TaxID=9548;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=98250422; PubMed=9590540;
 RA Jethanandani P., Santhanam R., Gupta S.K.;
 RT "Molecular cloning and expression in Escherichia coli of cDNA encoding bonnet monkey (Macaca radiata) zonae pellucida glycoprotein ZP2."
 RL Mol. Reprod. Dev. 50:229-239(1998).
 CC -!- FUNCTION: ZP2 FORMS WITH ZP1 AND ZP3 THE ZONAE PELLUCIDA. IN WHICH ZP2 AND ZP3 COMPLEX INTO COPOLYMERS CROSS-LINKED BY ZP1.
 CC ZP2 ACTS AS A SECONDARY SPERM RECEPTOR.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Extracellular matrix.
 CC -!- PTM: IS PROTEOLYTICALLY CLEAVED AFTER FERTILIZATION, AND THIS MODIFICATION, ALONG WITH PRESUMED CHANGES IN ZP3 MAY PLAY AN IMPORTANT ROLE IN THE POSTFERTILIZATION BLOCK TO POLYSPERMY.
 CC -!- PTM: SULFATED GLYCOPROTEIN WITH O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
 CC -!- SIMILARITY: Contains 1 ZP domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Y10690; CAA71693.1; -;
 DR InterPro; IPR001507; Endoglin/CD105.
 DR Pfam; PF00100; zona_pellucida; 1.
 DR PRINTS; PR00023; zPELLUCIDA.
 DR SMART; SM00241; ZP; 1.
 DR PROSITE; PS00682; ZP_DOMAIN; 1.
 KW Glycoprotein; Signal; Sulfate; Receptor; Transmembrane; Extracellular matrix.

FT SIGNAL 1 38
 FT CHAIN 39 745
 FT DOMAIN 39 716
 FT TRANSMEM 717 736
 FT DOMAIN 737 745
 FT DOMAIN 370 637
 FT CARBOHYD 87 87
 FT CARBOHYD 105 105
 FT CARBOHYD 122 122
 FT CARBOHYD 223 223
 FT CARBOHYD 269 269
 FT CARBOHYD 310 310
 FT CARBOHYD 400 400
 SQ SEQUENCE 745 AA; 82710 MW; 731D9AFA4D3EE028 CRC64;
 Query Match 33.5%; Score 45.5; DB 1; Length 745;
 Best Local Similarity 29.7%; Pred. NO. 57;
 Matches 11; Conservative 6; Mismatches 11; Indels 9; Gaps 1;
 QY 1 YVFSTEMA-----NKAEEAVLKQGVETIVSFH 28
 I : : : : : | : : : : : | : : : : :
 DB 382 VYSYQTQPALDLRLVGNSSCQPVFKAQSGQLVRFH 418
 RESULT 13
 PEM3_ARATH STANDARD; PRT; 490 AA.
 ID PEM3_ARATH
 AC Q9C6B9; Q9C9V1;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Putative phosphoethanolamine N-methyltransferase 3 (EC 2.1.1.103).
 GN NMT3 OR ARLG73600 OR F6D5.1 OR F5P22.1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y., Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W., Cressy T.H., Dewar K., Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y., Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L., Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E., Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B., Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P., Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A., Miltscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I., Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D., Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M., Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D., Uitterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M., Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana."
 RL Nature 408:816-820(2000).
 RN [2]
 RP CONCEPTUAL TRANSLATION.
 RP Schneider M.;
 RA Unpublished observations (MAY-2002).
 CC -!- FUNCTION: Catalyzes N-methylation of phosphoethanolamine, phosphomonomethyl ethanolamine and phosphodimethyl ethanolamine, the three methylation steps required to convert phosphoethanolamine to phosphocholine (By similarity).
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + ethanolamine phosphate = S-adenosyl-L-homocysteine + N-methyl ethanolamine phosphate.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:33 ; Search time 16.2335 Seconds
(without alignments)
445.097 Million cell updates/sec

Title: US-09-915-543-15_copy_177_204
Perfect score: 136
Sequence: 1 VYVFSTEMANKAAEAVLKGQVETIVSPH 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	40.4	1050	Q9ULQ4	Q9ulq4 caenorhabdi
2	51	37.5	243	Q97JA0	Q97ja0 clostridium
3	51	37.5	1049	Q9V6L1	Q9v6l1 drosophila
4	51	37.5	1049	Q960E6	Q960e6 drosophila
5	50	36.8	459	Q82521	Q82521 capsicum ch
6	50	36.8	642	Q9H6R2	Q9h6r2 homo sapien
7	50	36.8	642	Q96GN2	Q96gn2 homo sapien
8	50	36.8	1098	Q96Q02	Q96q02 homo sapien
9	49.5	36.4	350	Q8GGI9	Q8ggi9 lactobacill
10	49	36.0	141	Q9LEW7	Q9lfe7 arabidopsis
11	49	36.0	143	Q976C9	Q976c9 sulfolobus
12	49	36.0	187	Q8G3S3	Q8g3s3 bifidobacte
13	49	36.0	210	Q9F2L1	Q9f2l1 arabidopsis
14	49	36.0	233	Q96VA9	Q96va9 sulfolobus
15	49	36.0	237	Q96X55	Q96x55 sulfolobus
16	49	36.0	268	Q8DFX2	Q8dfx2 vibrio vuln

17	49	36.0	458	16	Q83532	O83532 treponema p
18	49	36.0	586	16	Q9KMS4	Q9kms4 vibrio chol
19	49	36.0	1047	3	Q8TFZ1	Q8tfz1 aspergillus
20	48.5	35.7	311	16	Q9PIC9	Q9pic9 campylobact
21	48.5	35.7	688	5	O96497	O96497 trypanosoma
22	48	35.3	225	17	Q96Z50	Q96z50 sulfolobus
23	48	35.3	235	17	Q976P7	Q976p7 sulfolobus
24	48	35.3	330	16	Q9K7G4	Q9k7g4 bacillus ha
25	48	35.3	609	10	Q9LV35	Q9lv35 arabidopsis
26	48	35.3	639	11	Q8BIN9	Q8bin9 mus musculus
27	48	35.3	1034	3	Q9Y7E9	Q9y7e9 candida gla
28	48	35.3	1659	4	Q8IWO9	Q8iwo9 homo sapien
29	47.5	34.9	444	16	O8XXK5	O8xxk5 clostridium
30	47	34.6	161	2	Q8GJ63	Q8gj63 bartonella
31	47	34.6	319	10	O8I460	O8i460 arabidopsis
32	47	34.6	352	17	Q980F9	Q980f9 sulfolobus
33	47	34.6	399	16	Q8PIN0	Q8pin0 streptococc
34	47	34.6	476	10	Q8RVV4	Q8rvv4 lycopersico
35	47	34.6	581	11	Q922Q3	Q922q3 mus musculu
36	47	34.6	662	16	Q8Z7U3	Q8z7u3 salmonella
37	47	34.6	732	16	Q8EA14	Q8ea14 shewanella
38	47	34.6	1072	3	Q8WZK5	Q8wzx5 neurospora
39	47	34.6	1261	5	Q8MMT3	Q8mmt3 dictyosteli
40	47	34.6	2388	5	O8MYG5	O8myg5 dictyosteli
41	47	34.6	2473	12	O7I209	O7i209 grapevine 1
42	46.5	34.2	541	16	Q8EPD3	Q8epd3 shewanella
43	46.5	34.2	614	5	Q22551	Q22551 caenorhabdi
44	46.5	34.2	771	5	Q19380	Q19380 caenorhabdi
45	46	33.8	320	5	Q8IJX6	Q8ijx6 plasmodium

ALIGNMENTS

RESULT 1

Q9ULQ4 PRELIMINARY; PRT; 1050 AA.

AC Q9ULQ4;
 DT 01-MAY-2000 (TREMREL. 13, Created)
 DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
 DT 01-MAR-2003 (TREMREL. 23, Last annotation update)
 DE Y87G2A.5 protein.
 GN Y87G2A.5
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA White S.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DDJF databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT *Genome sequence of the nematode C.elegans: A platform for investigating biology.*
 RL Science 282:2012-2018(1998).
 DR EMBL; AL110500; CAB60428.1; -.
 DR HSSP: P96142; IGAX.
 DR WormPep; Y87G2A.5; CE24685.
 DR InterPro; IPR002300; trna-synt_la.
 DR InterPro; IPR001412; trna-synt_i.
 DR InterPro; IPR002303; trna-synt_val.
 DR Pfam; PF00133; trna-synt_1; 1.
 DR PRINTS; PR00986; TRNASYNTHAL.
 DR TIGRFAMs; TIGR00422; vals; 1.
 DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
 SQ SEQUENCE 1050 AA; 118920 MW; F33DB53587EAC057 CRC64;

Query Match 40.4%; Score 55; DB 5; Length 1050;
 Best Local Similarity 44.4%; Pred. No. 22;
 Matches 12; Conservative 2; Mismatches 13; Indels 0; Gaps 0;


```

RL Plant Physiol. 118:1102-1102(1998).
DR EMBL: AF085149; AAC78480.1; -.
DR HSPSP; P04181; 20AT.
DR InterPro; IPR005814; Aminotrans_3.
DR Pfam; PF00202; aminotran_3; 1.
DR PROSITE; PS00600; AA_TRANSFER_CLASS_3; 1.
KW Aminotransferase; Transferase.
SQ SEQUENCE 459 AA; 50729 MW; 02ABB4D728B524E4 CRC64;

Query Match 36.8%; Score 50; DB 10; Length 459;
Best Local Similarity 41.7%; Pred. No. 51;
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKQGVETIVSF 27
Db 196 FSTRANNLESILKKEGPETVAAF 219

RESULT 6
Q9H6R2 PRELIMINARY; PRT; 642 AA.
ID Q9H6R2 AC Q9H6R2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ21965.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hikiiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA NAKANO T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK025618; BABL5191.1; -.
DR HSPSP; P96142; LGAX.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR002303; tRNA-synt_val.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00986; TRNASYNTHAL.
KW Hypothetical protein.
SQ SEQUENCE 642 AA; 71578 MW; C9E37E1D742B7F1 CRC64;

Query Match 36.8%; Score 50; DB 4; Length 642;
Best Local Similarity 44.4%; Pred. No. 73;
Matches 12; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 2 YVFSTEMANKAAEAVLKQGVETIVSFH 28
Db 60 FVRCQEMCARAAKAVESCALELSPSFH 86

RESULT 7
Q96GN2 PRELIMINARY; PRT; 642 AA.
ID Q96GN2 AC Q96GN2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Straussberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC009355; AAH09355.1; -.

```


R A Pridmore R.D., Arigoni F.;
R T "the genome sequence of Bifidobacterium longum reflects its adaptation
R T to the human gastrointestinal tract.";
R L Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
RL EMBL; AE014802; AAC25468.1; -.
K W Glycosyltransferase; Transferase; Complete proteome.
S Q SEQUENCE 187 AA; 20617 MW; E697C3C12727DB1 CRC64;

Query Sequence 36.0%; Score 49; DB 16; Length 187;
Best Local Similarity 45.2%; Pred. No. 26;
Matches 14; Conservative 3; Mismatches 4; Indels 10; Gaps 1;

QY 7 EMANKAAE-----AVLKGOVETIVSF 27
||| ||| |:
Db 26 EMAALASEDYRDKNPLLVAVLGAVNTLVAF 56

RESULT 13

Q9FZLL PRELIMITARY; PRY; 210 AA.

ID Q9FZLL AC Q9FZLI; 16, Created)

DT DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)

DE F17L21.2.

OS Arabidopsis thaliana (Mouse-eart cross).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatopsida II; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OX eurosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
RN [1]

RP SEQUENCE FROM N.A.

RA Johnson-Hopson C., Brooks S., Buehler E., Chao Q., Khan S., Kim C.,
RA Shinn P., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn I.,
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaverl A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Eckert J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F17L21 from chromosome
I";
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.

RN [2]
RP SEQUENCE FROM N.A.

RA Eckert J.R.;
RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.

[3]
RP SEQUENCE FROM N.A.

RA Eckert J.R.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.

RN [4]
RP SEQUENCE FROM N.A.

RA Eckert J.R.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.

[5]
RP SEQUENCE FROM N.A.

RA Eckert J.R.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.

[6]
RP SEQUENCE FROM N.A.

RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaverl A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Eckert J.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.

DR EMBL; ACC004557; AAF99722.1; -.
InterPro: IPRO03822; PAH.
Pfam: PF02671; PAH; 1.
SEQUENCE 210 AA; 24697 MW; FCD8130CD75700AO CRC64;


```
Query Match          36.0%; Score 49; DB 10; Length 210;
Best Local Similarity 41.7%; Pred. No. 30;
Matches 10; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKQGVETIVSF 27
DB 181 WSRSTNKAADRLAKGELENNVTF 204
      :|      |||||: : ||::| |::|
      :|      |||||: : ||::| |::|

RESULT 14
Q96YA9 PRELIMINARY; PRT; 233 AA.
AC Q96YA9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical protein ST2259.
GN ST2259.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kwarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000989; BAB67368.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 233 AA; 27253 MW; 8B9E6FDB60EE36D CRC64;

Query Match          36.0%; Score 49; DB 17; Length 233;
Best Local Similarity 39.3%; Pred. No. 34;
Matches 11; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKQGVETIVSPH 28
DB 120 VLVLTKNVEKEAEKVLKTRIDKVVYIH 147
      | | | | | | | | | | | | | | | | | |
      | | | | | | | | | | | | | | | | | |

RESULT 15
Q96X55 PRELIMINARY; PRT; 237 AA.
AC Q96X55;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein ST1165.
GN ST1165 OR ST0857.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kwarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 119.268 Seconds
(without alignments)
258.182 Million cell updates/sec

Title: US-09-915-543-15_COPY_199_392

Perfect score: 1028

Sequence: 1 TIVSFHIONISNKKTERSTA.....LPDEKEFTQAQSGGPOQNP 194

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

- 1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1028	100.0	1426	23	Human legless homo
2	1028	100.0	1435	22	Human BCL9 homolog
3	277	26.9	140	22	Human polypeptide
4	232.5	22.6	320	23	Mouse beta-catenin
5	232.5	22.6	1494	23	Mouse beta-catenin
6	167.5	16.3	707	22	Human polypeptide
7	162.5	15.8	574	22	WASP homolog prote
8	162.5	15.8	574	22	Amino acid sequenc
9	159	15.5	406	22	Novel human diagno

10	158.5	15.4	256	22	ABU52945	Human mammary carc
11	158.5	15.4	270	22	ABU52939	Human mammary carc
12	158.5	15.4	446	22	ABW70063	Drosophila melanog
13	158	15.4	783	19	AAW37151	Mouse neural Menat
14	158	15.4	787	19	AAW37152	Mouse neural Menat
15	158	15.4	802	19	AAW37153	Mouse neural Menat
16	158	15.4	802	22	AAW09139	Mammalian enabled
17	157	15.3	598	22	ABG14000	Novel human diagno
18	155	15.1	253	22	ABU52937	Human mammary carc
19	155	15.1	572	18	AAW31855	Mycobacterium tube
20	155	15.1	763	18	AAW31852	Mycobacterium tube
21	154.5	15.0	731	22	ABW74209	Protein encoded by
22	154.5	15.0	731	23	ABW93202	Herbicidally activ
23	153.5	14.9	177	22	ABU52946	Human mammary carc
24	151.5	14.7	441	22	ABW11413	Human extensin hom
25	151	14.7	693	23	ABW69529	Human polypeptide
26	150.5	14.6	533	16	AAW72483	Human H-2RIIBP. H
27	150.5	14.6	533	23	ABW79960	Retinoid C recepto
28	150.5	14.6	533	23	AAO19271	Human retinoid X r
29	150.5	14.6	533	23	AAO19282	Human retinoid X r
30	150.5	14.6	533	23	AAO19283	Human retinoid X r
31	150.5	14.6	533	23	AAO19284	Human retinoid X r
32	150.5	14.6	533	23	AAO19285	Human retinoid X r
33	150.5	14.6	533	23	AAO19286	Human retinoid X r
34	150.5	14.6	533	23	AAO19287	Human retinoid X r
35	150.5	14.6	533	23	AAO19288	Human retinoid X r
36	150.5	14.6	533	23	AAO19289	Human retinoid X r
37	150.5	14.6	533	23	AAO19291	Human retinoid X r
38	150.5	14.6	533	23	AAO19292	Human retinoid X r
39	150.5	14.6	533	23	AAO19293	Human retinoid X r
40	150.5	14.6	533	23	AAO19294	Human retinoid X r
41	150.5	14.6	533	23	AAO19295	Human retinoid X r
42	150.5	14.6	533	23	AAO19296	Human retinoid X r
43	150.5	14.6	533	23	AAO19297	Human retinoid X r
44	150.5	14.6	533	23	AAO19298	Human retinoid X r
45	150.5	14.6	533	23	AAO19299	Human retinoid X r

ALIGNMENTS

RESULT 1

AAW71229
ID AAW71229 standard; Protein; 1426 AA.

AC AAW71229;

DT 18-NOV-2002 (first entry)

DE Human legless homologue lgs/bcl9 protein.

KW Legless; human; lgs; Wnt/Wingless signaling pathway; Wnt; Wg; tissue proliferation; tumour; cytostatic; cellular disorder; colon; blood disorder; cancer; breast; head and neck cancer; brain; thyroid; medulloblastoma; skin cancer; tissue regeneration; tissue repair.

OS Homo sapiens.

PN US2002086986-A1.

PD 04-JUL-2002.

PF 27-JUL-2001; 2001US-0915543.

PR 28-JUL-2000; 2000US-221502P.

PA (BASL/) BASLER K.

PA (BRUN/) BRUNNER E.

PA (FROE/) FROESCH B.

PA (KRAM/) KRAMPS T.

PA (PETE/) PETER O.

PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;

```
XX DR WPI; 2002-635689/68.
XX DR N-PSDB; AAF88467.
XX
XX FT Novel polypeptide useful in therapeutic method for treating disorders
XX PT of cell fate such as cell differentiation or cell proliferation -
XX PS
XX PS Example II; Fig 8B; 41pp; English.
XX
XX CC This invention describes a novel polypeptide sharing one or more
XX CC homologous amino acid domains with the legless (lgs) protein, a
XX CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway
XX CC involved in the formation and maintenance of spatial arrangements
XX CC and proliferation of tissues during development, and in the formation
XX CC and growth of many human tumours. The products of the invention have
XX CC cytosolic activity and can be used to treat cellular disorders, blood
XX CC disorders and cancers caused by over-stimulation of the Wnt pathway,
XX CC where the cancerous condition is colon, breast, head and neck, brain,
XX CC thyroid, medulloblastoma or skin cancer. The product could also be used
XX CC to promote tissue regeneration and repair. This sequence represents the
XX CC human legless (lgs) protein homologue lgs/bcl9 described in the
XX CC disclosure of the invention.
XX
XX CC Sequence 1426 AA;
XX
XX CC Query Match 100.0%; Score 1028; DB 23; Length 1426;
XX CC Best Local Similarity 100.0%; Pred. No. 7.5e-65;
XX CC Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 TIVSFHQNISNNKTERSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNTRLQPTP 60
Db 199 TIVSFHQNISNNKTERSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNTRLQPTP 258
QY 61 PIPAPAPKPAAPRPLDRESFGVENKLIPIPSVGSPASSTPLPDGTGPNSTNNRAVTPVS 120
Db 259 PIPAPAPKPAAPRPLDRESFGVENKLIPIPSVGSPASSTPLPDGTGPNSTNNRAVTPVS 318
QY 121 QGSNSSADKPAPPPPPVSSGEPPTLGNPDGLSQEQLHRSLSQTLRDIQRMFLFPEDEK 180
Db 319 QGSNSSADKPAPPPPPVSSGEPPTLGNPDGLSQEQLHRSLSQTLRDIQRMFLFPEDEK 378
QY 181 EFTGAQSGGPOQNP 194
Db 379 EFTGAQSGGPOQNP 392
RESULT 2
AB111808
ID AB111808 standard; peptide; 1435 AA.
AC
XX
XX AC AB111808;
XX
XX DT 11-JAN-2002 (first entry)
XX
XX DE Human BCL9 homologue, SEQ ID NO:2178.
XX
XX KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
XX KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
XX KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
XX KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
XX KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
XX KW chronic inflammatory condition; proliferative retinopathy;
XX KW atherosclerosis; coronary heart disease; arterial ischaemia;
XX KW bone disorder; osteoporosis; vascular growth disorder;
XX KW tissue regeneration; wound healing; infection; immune disorder;
XX KW cell culture; drug screening; gene therapy; antiinflammatory;
XX KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
XX KW cytosolic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
XX KW antifungal; vulnery; antiulcer.
XX
XX OS Homo sapiens.
XX
XX XX WO200157188-A2.
XX
XX
```

```
XX 09-AUG-2001.
XX
XX PF 05-FEB-2001; 2001WO-US03800.
XX
XX PR 03-FEB-2000; 2000US-0496914.
XX PR 27-APR-2000; 2000US-0560875.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX PI Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-457740/49.
XX N-PSDB; ABA09052.
XX
XX PT Human proteins and DNA encoding sequences useful for preventing,
XX PT treating or ameliorating a medical condition in a mammalian subject
XX PT e.g. arthritis and cancer -
XX
XX PS Claim 20; Page 256-257; 1963pp; English.
XX
XX CC Sequences AB10981-AB12330 represent 1350 novel human polypeptides, and
XX CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
XX CC invention also relates to vectors and recombinant host cells comprising a
XX CC nucleotide of the invention, methods of producing the novel polypeptides,
XX CC antibodies against the polypeptides, methods of detecting the nucleotides
XX CC or polypeptides in a sample, and methods of identifying compounds which
XX CC bind to polypeptides of the invention. Although novel, many of the
XX CC polypeptides of the invention have homology to known proteins, and hereby
XX CC giving an insight into their probable biological activities, and hence
XX CC potential therapeutic applications. The polypeptides of the invention may
XX CC have various activities, including cytokine, cell proliferation or cell
XX CC differentiation activities; stem cell growth factor activity;
XX CC haematopoiesis regulatory activity; tissue growth activity;
XX CC immunomodulatory activity; activin- or inhibin-related activities;
XX CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
XX CC thrombolytic activities; receptor or ligand activities; or may be
XX CC involved in oncogenesis; cancer cell proliferation or metastasis.
XX CC Depending on their biological activities, polypeptides and nucleotides of
XX CC the invention are useful for preventing, treating or ameliorating medical
XX CC conditions, e.g., by protein or gene therapy. Such conditions include
XX CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
XX CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
XX CC proliferative retinopathy, atherosclerosis, coronary heart disease,
XX CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
XX CC vascular growth. Polypeptides involved with tissue regeneration and
XX CC repair (or nucleic acids encoding them) may be used to promote wound
XX CC healing (e.g., of burns, incisions and ulcers), while those with
XX CC immunomodulatory activities may be used in the treatment of viral,
XX CC bacterial and fungal infections in addition to immune disorders.
XX CC Polypeptides with growth factor activity may be used in cell cultures to
XX CC promote cell growth. For example, such polypeptides may be used to
XX CC manipulate stem cells in culture to give rise to neuroepithelial cells
XX CC that can be used to augment or replace cells damaged by illness,
XX CC autoimmune disease or accidental damage. The polypeptides and nucleotides
XX CC may also be used in the diagnosis of the above conditions, and in drug
XX CC screening techniques. The present sequence represents a novel human
XX CC polypeptide of the invention.
XX
XX CC Sequence 1435 AA;
XX
XX CC Query Match 100.0%; Score 1028; DB 22; Length 1435;
XX CC Best Local Similarity 100.0%; Pred. No. 7.6e-65;
XX CC Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 TIVSFHQNISNNKTERSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNTRLQPTP 60
Db 239 TIVSFHQNISNNKTERSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNTRLQPTP 298
QY 61 PIPAPAPKPAAPRPLDRESFGVENKLIPIPSVGSPASSTPLPDGTGPNSTNNRAVTPVS 120
Db 299 PIPAPAPKPAAPRPLDRESFGVENKLIPIPSVGSPASSTPLPDGTGPNSTNNRAVTPVS 358
```

RESULT 4
AAU78461
ID AAU78461 standard; Protein; 320 AA.

AA	Mouse beta-catenin nuclear localised protein #2.
DE	
XX	
XX	
KW	Mouse; beta-catenin nuclear localised protein; cancer;
KW	gene therapy; EST; expressed sequence tag.

KW	gene therapy
XX	
OS	Mus musculus
XX	
PN	WO200224738-A

OS Mus musculus.
XX
PN WO200224738-A1.
XX
PD 28-MAR-2002.

EN WO200224738-A1.
XX
XX
PD 28-MAR-2002.
XX
XX
PF 19-SEP-2001: 2001WO-JE08140.

PD	28-MAR-2002.
XX	
PF	19-SEP-2001;
XX	
PR	22-SEP-2000;

PF 19-SEP-2001; 2001WO-JF08140.
XX
PR 22-SEP-2000; 2000JP-0387876.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX

PR 22-SEP-2000;
XX
PA (KYOW) KYOW
XX
PI Akiyama T.

PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX
XX Akiyama T, Adachi S;
PI
XX
XX WPI: 2002-330014/36.
DR

PI Akiyama T,
XX
DR WPI; 2002-33
DR N-PSDB; ABK4
XX

DR
DR
DR
xx
PT
PT
PT

WPI: 2002-330014/36.
N-PSDB; ABK47632.

New beta-catenin nuclear localised protein for diagnosis of diseases associated with nuclear localisation of beta-catenin

XX	New beta-cat
PT	of diseases
PT	cancer
PT	-
XX	

PT of diseases associated with nuclear localisation of
 PT cancer -
 XX
 XX
 PS Claim 2; Page 91-92; 113pp; Japanese.
 XX

XX	Claim 2; pag
PS	
XX	The inventio
CC	and DNA encod
CC	

xx The invention relates to a beta-catenin nuclear localisation and DNA encoding the protein. The protein and encoding DNA are applicable in diagnosis and treatment of diseases associated with nuclear localisation of beta-catenin e.g. cancer. In particular, the invention relates to a beta-catenin nuclear localisation and DNA encoding the protein.

CC and DNA encod
CC applicable i
CC nuclear loca
CC therapy. The
CC mouse beta-c

CC nuclear localisation of beta-catenin e.g. cancer, incl
CC therapy. The present sequence represents the amino acid
CC mouse beta-catenin nuclear localised protein #2.
XX
SO Sequence 320 AA:

```

CC mouse beta-c
XX
SQ Sequence 3
Query Match

```

SQ	Sequence	320 AA;
Query Match	22.6%	Score 232.5; DB 23; Len
Best Local Similarity	35.1%	Pred. No. 5.7e-09; Len
Matches	71; Conservatives	23; Mismatches 775; Index

Query Match
Best Local Similarity
Matches 71; 0
Qv 1 TIV

	Matches	71;	Conservative	23;	Mismatches	75;	Indels
Qy	1	TIVSHIQINISNNKTERSTAPLNTQISALRNDPKLPQQPPAPPANPANN					
Dd	16	SILAYHQHNVPRAKLDQA-----KKVPTTDEPLPLN					
		: : : : : :					

Qy	1 TIV : :
Db	16 SIL
Qy	61 PIP

[illegible]

QY	61 PIP
	:
Db	59 PLP
QY	118 PVS

D6 DB QY D6 DB

59 PLPPPPPAPGSADPALPPEGGPEDTSDQLAPNSVG--AASTGGGCT
118 PVSQGSNSSSADPKA--PPPPVSSGEPPTLGENPDFLSQEQLHEHH
! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! :
117 PLPPGGDPGSAPGSALGEATGTNGRNILVGS EGLSKQEQLFEEHHE

QY	118 PVS
	:
Db	117 PLP
QY	176 F--

```

117 PLPPGDDPGSAPGSALLCEATPTGNGORNVLGS--EGLSKQEQLHEH
QY 176 F--PDEKEFT-----GAQSGGP 190
      : : | |||
175 LRSGETEPFLKPGPGGAGEGG 196
Db

```

QY	176 F--
Db	175 LRS

DB 175 LRSGETEPLKGGPGGAGEGP 196

RESULT 5
AAU78460

RESULT 5
AAU78460
ID AAU78460 sta
XX

AAU78460
ID AAU78460 standard; Protein: 1494 AA.
XX
AC AAU78460;
XX

XX AC XX
AAU78460;


```

XX KW Human; gene therapy; vaccine; disease treatment; detection.
XX OS Homo sapiens.
XX PN WO200112659-A2.
XX PD 22-FEB-2001.
XX XX
XX XX 18-AUG-2000; 2000WO-IB01496.
XX PF 18-AUG-1999; 99US-0149499.
XX PR 28-SEP-1999; 99US-0156503.
XX XX
XX XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX PA Wiemann S;
XX PI WPI; 2001-327840/34.
XX DR
XX XX
XX PT Nucleic acids having the sequences of clones isolated from libraries of
XX PT different human tissues, useful in recombinant DNA methodologies -
XX PS Example III; Page 550; 1095pp; English.
XX XX
XX CC This invention describes novel polynucleotides and polypeptides isolated
XX CC from human cDNA libraries which can be used for gene therapy or in
XX CC vaccines. The polynucleotides of the invention and antibodies encoded by
XX CC them may be used in the prevention, diagnosis and treatment of diseases
XX CC associated with inappropriate polypeptide expression. The products of the
XX CC invention may also be used to identify modulators of expression and
XX CC activity and to down regulate expression and activity. The antibodies of
XX CC the invention may also be used as diagnostic agents for detecting the
XX CC presence of polypeptides in samples. This sequence represents a homologue
XX CC of a polypeptide described in the disclosure of the invention.
XX SQ Sequence 270 AA;

Query Match 15.4%; Score 158.5; DB 22; Length 270;
Best Local Similarity 32.0%; Pred. No. 0.00087;
Matches 49; Conservative 17; Mismatches 68; Indels 19; Gaps 7;

QY 14 KTERSTAPLNTQISALRNDPKLP--QQPPAPANQDNSSQNTRLQP---TPPIAPAPK 68
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
79 KSSPPAPVNLPPPEVKSSPPPTPVSSPPAPKSSPPPPAPMSSPPPEVKSSPPPPAPVSS 138
QY 69 P-----AAPRPLDRESGVENKLIPIVSGSPASSTPLP---PDGTGPNSTNNRAVTPV 119
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
139 PPPPVKSSPPPPAPVSSPPPPVKS---PPPPAPVSSPPPPVKSPPPPAPISPPPPVKSPP 195
QY 120 SQGSNSSADP-KAPPPP-PVSSGEPPTLGENP 150
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
196 PPAPVSSPPPPVKSPPPPAPVSSPPPPVKSPPPP 228

RESULT 12
ABB70063
ID ABB70063 standard; Protein; 446 AA.
XX AC ABB70063;
XX XX
XX DT 26-MAR-2002 (first entry)
XX DE
XX DE Drosophila melanogaster polypeptide SEQ ID NO 36981.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO2001171042-A2.
XX PD 27-SEP-2001.

```

```

XX 23-MAR-2001; 2001WO-US09231.
XX PF
XX XX
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX XX
XX PA (PEKE ) PE CORP NY.
XX PI
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX XX
XX WPI; 2001-656860/75.
XX DR N-PSDB; ABL14166.
XX XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX XX
XX PS Disclosure; SEQ ID NO 36981; 21pp + Sequence Listing; English.
XX XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (AB95737-AB972072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 446 AA;

Query Match 15.4%; Score 158.5; DB 22; Length 446;
Best Local Similarity 29.6%; Pred. No. 0.0014;
Matches 42; Conservative 11; Mismatches 64; Indels 25; Gaps 3;

QY 33 PKPLPQQPPAPANQDNSSQNTRLQTPPIPAPA--PKPAAPPRLDRESGVENKLIPI 90
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
97 PQTTPPPPPPP-----QTPPAPRPSYGPQTQPPRPPPTTPSAPAPPPPS 144
QY 91 VG-----SPASSTPLPDDGTGPNSTNNRAVTPVSGSNSSADPKAPPPPV 139
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
145 YGPPQTTPRPPPPQTPPSAPAPPPSYGPPQTTPRPPPPQTPSAPAPSYGPPQPPAPQ 204
QY 140 SGEPTLGENPDGLSQEQLEHR 161
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
205 PPSPPSPQGPETLPPDPQKPR 226

RESULT 13
AAW37151
ID AAW37151 standard; Protein; 783 AA.
XX AC AAW37151;
XX XX
XX DT 06-JUL-1998 (first entry)
XX DE
XX DE Mouse neural Mena+ protein.
XX XX
XX KW Neural Mena+ protein; mammalian Ena; Enabled protein; Evi protein;
XX KW cytoskeleton; cell morphology; cell adhesion; cell differentiation;
XX KW cell growth; cell motility; mouse.
XX XX
XX OS Mus musculus.
XX XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 378
XX XX /note= 'encoded by GGN'
XX XX
XX PN WO9801755-A1.
XX XX
XX PD 15-JAN-1998.

```


KW Neural Mena⁺⁺ protein; mammalian Ena; Enabled protein; Evi protein;
 KW cytoskeleton; cell morphology; cell adhesion; cell differentiation;
 KW cell growth; cell motility; mouse.

OS Mus musculus.

AA WO9801755-A1.
PN

15-JAN-1998.

XX 03-JUL-1997; 97WO-US11669.
PF

XX PR 05-JUL-1996; 96US-0675815.

PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.

PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
 (GDB) GES BIOLOGISCHE FORSCHUNG

PI Gertler FB, Niebuhr K, Soriano P, Wehland J;

DR WPI; 1998-101197/09.

AA Detection of modulators of Mena and Ena-VASP-like genes and proteins
 PT - used in control of cytoskeletal dynamic events in normal and
 PT abnormal cell morphology, adhesion, motility, growth and
 PT differentiation

PS Example 4; Page 63-65; 77pp; English.

This protein comprises novel murine neural Mena⁺⁺⁺. Its amino acid sequence was deduced from a cDNA clone obtained from a mouse brain cDNA library. Two other isoforms, neural Mena⁺ (see AAW37151) and neural Mena⁺⁺ (see AAW37152), are also disclosed. Unlike mammalian Ena (Mena, see AAW37148), neural Mena isoforms exhibit neural tissue-specific distribution. Based on the disclosed Mena and Evl genes (see also AAW02996-98) and proteins (see also AAW37148-49), a variety of methods and compositions are provided for screening, isolating and characterising endogenous and exogenous factors, drugs and therapeutic agents useful to evaluate and/or control cytoskeletal dynamic events involved in normal and abnormal cell morphology, adhesion, motility, growth and/or differentiation. A method of detecting a modulator of Mena activity/expression is claimed.

Sequence 802 AA;

Query Match 15.48; Score 158; DB 19; Length 802;

Best Local Similarity 26.3%; Pred. No. 0.0028;
 Exact match 15.4%; Score 158; DB 13; Length 802;

2002 Local Similarity 20.56; P-Value: 0.0028;
 Matches 59; Conservative 15; Mismatches 62; Indels 88; Gaps 10;

QY	13 NKTERSTAPLNT-----QISALRNDP-----	33
----	------------------------------------	----

[illegible]

08. CACAPITA INDEUTENS, I JOCTIVT I AC Y TE TC TINOY INUROCONHTT I TNOO TANA TOO INEQUENT 0.00

QY 34 KPLPQQPPAPANQDN-----SSQNTRLQPTPIPAPAPKPAAPRPL----- 76

Db

408 SPVPQMPSPPTAPNGSLDSVTYPVSPPTSGPAAAPPPPPPPPPPPPPPLAS 467

[illegible]

QY ---DRESFGVFNKLI---PSVGSPASS-TPLPPD-GTGFNSTPNNRAV-ILG

// : | | : | | : | | : | | : | | : | | :

Db 468 LSHCGSQASPPPGTPLASTPSSKPSVLPSPSACAPASAEPLNPELGDSSASEPGLQAAS 527

Search completed: October 15, 2003, 10:30:25
Job time : 120.268 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 39.2529 Seconds
(without alignments)
209.113 Million cell updates/sec

Title: US-09-915-543-15_COPY_199_392

Perfect score: 1028
Sequence: 1 TIVSFHIONISNNKTERSTA.....LFPDEKEFTGAQGGPQQNP 194

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2.6/protdata/1/iaa/5A_COMB.pep:*
2: /cgn2.6/protdata/1/iaa/5B_COMB.pep:*
3: /cgn2.6/protdata/1/iaa/6A_COMB.pep:*
4: /cgn2.6/protdata/1/iaa/6B_COMB.pep:*
5: /cgn2.6/protdata/1/iaa/6C_COMB.pep:*
6: /cgn2.6/protdata/1/iaa/6D_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	161.5	15.7	816	4	US-09-266-225D-12
2	150.5	14.6	533	1	US-07-952-800-2
3	149.5	14.5	525	3	US-08-764-870-7
4	149.5	14.5	525	3	US-08-980-115-7
5	143.5	14.0	446	1	US-07-952-800-4
6	140.5	13.7	8991	4	US-08-714-741-32
7	138.5	13.5	297	2	US-08-580-5458-6
8	138.5	13.5	297	3	US-09-262-653A-6
9	137	13.3	456	2	US-08-819-013-1
10	136.5	13.3	142	4	US-09-252-991A-24873
11	136.5	13.3	1248	2	US-09-080-897-2
12	136.5	13.3	1248	3	US-09-323-735-2
13	135	13.1	332	3	US-08-818-112-53
14	135	13.1	332	4	US-08-818-111-53
15	135	13.1	332	4	US-09-056-556-53
16	135	13.1	332	4	US-09-072-596-53
17	135	13.1	334	6	5202236-3
18	133.5	13.0	941	4	US-07-757-022B-14
19	133.5	13.0	1022	4	US-07-757-022B-84
20	133.5	13.0	1038	4	US-07-757-022B-74
21	133.5	13.0	1049	4	US-07-757-022B-58
22	133.5	13.0	1140	4	US-07-757-022B-104
23	133.5	13.0	1270	4	US-07-757-022B-44
24	133.5	13.0	1311	4	US-07-757-022B-42
25	133.5	13.0	1313	4	US-07-757-022B-142
26	133.5	13.0	1314	4	US-07-757-022B-50
27	133.5	13.0	1320	4	US-07-757-022B-46

28	133.5	13.0	1320	4	US-07-757-022B-60	Sequence 60, Appl
29	133.5	13.0	1354	4	US-07-757-022B-48	Sequence 48, Appl
30	133.5	13.0	1361	4	US-07-757-022B-40	Sequence 40, Appl
31	133.5	13.0	1363	4	US-07-757-022B-52	Sequence 52, Appl
32	133.5	13.0	1404	4	US-07-757-022B-2	Sequence 2, Appl
33	133.5	13.0	1404	4	US-07-757-022B-62	Sequence 62, Appl
34	133	12.9	1315	3	US-08-899-595-3	Sequence 3, Appl
35	132	12.8	214	1	US-08-217-327-4	Sequence 4, Appl
36	132	12.8	331	6	5202236-37	Patent No. 5202236
37	132	12.8	1231	4	US-08-714-741-41	Sequence 41, Appl
38	131	12.7	2442	4	US-09-514-247A-10	Sequence 10, Appl
39	130.5	12.7	325	1	US-08-382-184-2	Sequence 2, Appl
40	130.5	12.7	325	3	US-08-641-356-2	Sequence 2, Appl
41	130.5	12.7	325	3	US-09-132-528-2	Sequence 2, Appl
42	130.5	12.7	325	3	US-09-132-528-3	Sequence 3, Appl
43	130.5	12.7	325	3	US-08-875-494-2	Sequence 2, Appl
44	130.5	12.7	325	4	US-09-599-366-2	Sequence 2, Appl
45	130.5	12.7	325	4	US-09-599-366-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-266-225D-12
; Sequence 12, Application US/09266225D
; Patent No. 6573364
; GENERAL INFORMATION:
; APPLICANT: Nandabalan, Krishan
; APPLICANT: Kingsmore, Stephen
; APPLICANT: Tchernev, Velizar
; TITLE OF INVENTION: Isolation and Characterization of Hermansky-Pudlak
; TITLE OF INVENTION: Syndrome (HPS) Protein Complexes and HPS Protein-
; TITLE OF INVENTION: Interacting Proteins
; FILE REFERENCE: 15966-523
; CURRENT APPLICATION NUMBER: US/09/266,225D
; CURRENT FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 12
; LENGTH: 816
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-266-225D-12

Query Match 15.7%; Score 161.5; DB 4; Length 816;
Best Local Similarity 32.1%; Pred. No. 1.6e-05;
Matches 54; Conservative 7; Mismatches 64; Indels 43; Gaps 7;
QY 17 RSTAPLNTQISALRNDPKPLP-----QQPPAPANQDQNSSQNTRIQPTPTIP 63
Db 577 RPAAPALTSVPAPAPAPATPTPTVQPTSPPPGVAQPTGTPQFSAGSTSGPVPPQACPPP 636
QY 64 APAPKPAAPRPRLDRESPG-----VENKLIPSGVSPASSTP-----LPPDGT 105
Db 637 GRAPHTGTPGPIPVPPAPQIATSTLLAAQSLVPPPLPGSTPGVLVFPFPLPPDA 696
QY 106 G--PNS-----TPNNRAVTPVSGNSSSADKAPPPPVSSGEPFLG 147
Db 697 GGAPQSSMESPDVNLVT--QQLSKSQVEDPL----PPVFSGTPKSG 738

RESULT 2
US-07-952-800-2
; Sequence 2, Application US/07952800
; Patent No. 5403925
; GENERAL INFORMATION:
; APPLICANT: OKATO, KEIKO
; TITLE OF INVENTION: A NEW MEMBER OF THE NUCLEAR HORMONE
; TITLE OF INVENTION: RECEPTOR SUPERFAMILY AND A CDNA CLONE THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND AND TOWNSEND

STREET: 1 MARKET PLAZA, STEUART TOWER, SUITE 2000
CITY: SAN FRANCISCO
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/952,800
FILING DATE: 19920928
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-21-1
TELEPHONE: 415-543-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 533 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-952-800-2

Query Match 14.6%; Score 150.5; DB 1; Length 533;
Best Local Similarity 31.0%; Pred. No. 7.7e-05;
Matches 40; Conservative 11; Mismatches 45; Indels 33; Gaps 5;

QY 31 NDPKLPQ--QPPANQDNSSQNRTRQPTPIAPA-----PKFAAPRPLDRSP 81
Db : | | | | | : | | | | : | | | | : | | | | : | | | | :
86 SSPNLPQGVPPSPG-----PPLPSTAPSLGGSGAPPPMPPLGSPFP 134
QY 82 GVENKLIPVSGSPASTPLPPDGTGPNSTPNNRATVPVSGNSSADPKAP----- 133
Db : | | | | | : | | | | : | | | | : | | | | : | | | | :
135 VISS-----SMGSPGLPPAPPFGVSPVSPQINSVSLPFGGSGPPEDVKPVLGVRLH 190
QY 134 -PPPVSSG 141
Db : | | | | :
191 CPPPGGPG 199

RESULT 3
US-08-764-870-7
Sequence 7, Application US/08764870
Patent No. 6236946
GENERAL INFORMATION:
APPLICANT: Scanlan, Thomas S
APPLICANT: Baxter, John D
APPLICANT: Fletterick, Robert J
APPLICANT: Wagner, Richard L
APPLICANT: Kushner, Peter J
APPLICANT: Apriletti, James W
APPLICANT: West, Brian
TITLE OF INVENTION: Nuclear Receptor Ligands and Ligand
TITLE OF INVENTION: Binding Domains
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooley Godward
STREET: Five Palo Alto Square, 3000 El Camino Real
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,870
FILING DATE: 13-DEC-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/008,540
FILING DATE: 13-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/008,543
FILING DATE: 13-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/008,606
FILING DATE: 14-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Nakamura, Jackie N
REGISTRATION NUMBER: 35,966
REFERENCE/DOCKET NUMBER: UCAL-246/010US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)843-5000
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 525 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-764-870-7

Query Match 14.5%; Score 149.5; DB 3; Length 525;
Best Local Similarity 31.0%; Pred. No. 9.1e-05;
Matches 40; Conservative 10; Mismatches 46; Indels 33; Gaps 5;

QY 31 NDPKLPQ--QPPANQDNSSQNRTRQPTPIAPA-----PKFAAPRPLDRSP 81
Db : | | | | | : | | | | : | | | | : | | | | : | | | | :
78 SSPNLPQGVPPSPG-----PPLPSTAPTLGGSGAPPPMPPLGSPFP 126
QY 82 GVENKLIPVSGSPASTPLPPDGTGPNSTPNNRATVPVSGNSSADPKAP----- 133
Db : | | | | | : | | | | : | | | | : | | | | : | | | | :
127 VISS-----SMGSPGLPPAPPFGVSPVSPQINSVSLPFGGSGPPEDVKPVLGVRLH 182
QY 134 -PPPVSSG 141
Db : | | | | :
183 CPPPGGPG 191

RESULT 4
US-08-980-115-7
Sequence 7, Application US/08980115
Patent No. 6266622
GENERAL INFORMATION:
APPLICANT: Scanlan, Thomas S
APPLICANT: Baxter, John D
APPLICANT: Fletterick, Robert J
APPLICANT: Wagner, Richard L
APPLICANT: Kushner, Peter J
APPLICANT: Apriletti, James W
APPLICANT: West, Brian L
APPLICANT: Shiao, Andrew K
TITLE OF INVENTION: NUCLEAR RECEPTOR LIGANDS AND LIGAND BINDING DOMAINS
FILE REFERENCE: UCAL-246/020S
CURRENT APPLICATION NUMBER: US/08/980,115
CURRENT FILING DATE: 1997-11-26
EARLIER APPLICATION NUMBER: 08/764,870
EARLIER FILING DATE: 1996-12-13
EARLIER APPLICATION NUMBER: 60/008,606
EARLIER FILING DATE: 1995-12-14
EARLIER APPLICATION NUMBER: 60/008,543
EARLIER FILING DATE: 1995-12-13
EARLIER APPLICATION NUMBER: 60/008,540
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7


```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/788,322
; FILING DATE: 24-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-64383-1/RET/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-819-013-1

Query Match 13.3%; Score 137; DB 2; Length 456;
Best Local Similarity 25.9%; Pred. No. 0.00083;
Matches 51; Conservative 26; Mismatches 58; Indels 62; Gaps 10;

QY 10 ISNNKTERSTAPLNTQISALRNDPKPLPQQPPAPANQDQNSQNTLRLOTPPIAPAPAKP 69
DB 120 IDNRSSQRHSPFS-----KTLPSKPSWP-----SEKARLSTLPALTALQKP 162
QY 70 AAPRP-----LDRESPV-----ENKLIPSVGS--PASSTPLPDGPGNTPNNRAVT 117
DB 163 QVPPKPKGLLEADYVYVVEDNDENYIHTPESSPPPEKAPMVNRSTKPNSS-----T 216
QY 118 PVS-----QGSNSSADPKAPP-----PPVSSGEPPTL-----GENPDGLSQEQ-- 157
DB 217 PASPPGTASGRNSGAWETKSPPAAPSLPRAGKAPTTLKTTVPASQONASSVCEKPI 276
QY 158 -----LEHRSLSQT 167
DB 277 PAERHRGSSHQEAQVS 293

RESULT 10
US-09-252-991A-24873
; Sequence 24873, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24873
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24873

Query Match 13.3%; Score 136.5; DB 4; Length 142;
Best Local Similarity 26.8%; Pred. No. 0.00023;
Matches 41; Conservative 19; Mismatches 36; Indels 57; Gaps 7;

QY 7 IONISNNKTERSTAPL-----NTQISALRNDPKPLPQQPPAPANQDQNSQNTLRLOPT 59
DB 25 VQATCSDSKSGCTLSLNLGNLFTNAQSLLQPPCKPVP----- 62
QY 60 PPIAPAPKPAAPRPPLDRSPGVENKLIPSVGSPASSTPLPDPGTGPNSTPNNRAVTPV 119
```

```
DB 63 PPVPPSPPEPSPPP--PVPPSPPV-----PGVPPSPSPPPVP--SPPPSP----- 106
QY 120 SQGSNSSSADPKAPPPPPVSSGEPPTLGENPDG 152
DB 107 -----PPVPPSPPPS--PPPPVPPSPDG 127

RESULT 11
US-09-080-897-2
; Sequence 2, Application US/09080897
; Patent No. 5985574
; GENERAL INFORMATION:
; APPLICANT: King, Mary-Claire
; APPLICANT: Lynch, Eric D.
; APPLICANT: Lee, Ming
; APPLICANT: Morrow, Jan E.
; APPLICANT: Welcsh, Piri L.
; APPLICANT: Leon, Pedro E.
; TITLE OF INVENTION: Modulators of Actin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,897
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UW97-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1248 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-080-897-2

Query Match 13.3%; Score 136.5; DB 2; Length 1248;
Best Local Similarity 32.3%; Pred. No. 0.0029;
Matches 43; Conservative 8; Mismatches 59; Indels 23; Gaps 5;

QY 35 PLPQQ-----PPAPANQDQNSQNTLRLOTPPIAPAPAKPA-----APRPLDRS----- 80
DB 578 PLPGDSGTIIIPPPAPGD-----STTPPPPPPPPPPPPLPGGTALSPPLSGDATIPPP 633
QY 81 -PGVENKLIPSVGSPASSTPLPDPGTGPNSTPNNRAVTPVSGSNSSSADPKAPPPPPVS 139
DB 634 PPLPEGVCIPIPSLSLPGGTALPP-----PPPLPGSARIPPPPPPLPGSAGIPPPPPPLPGE 689
QY 140 SGEPTLGENPDG 152
DB 690 AGMPPPPPPLPGG 702

RESULT 12
US-09-323-735-2
; Sequence 2, Application US/09323735
; Patent No. 6197932
```

```

; GENERAL INFORMATION:
; APPLICANT: King, Mary-Claire
; APPLICANT: Lynch, Eric D.
; APPLICANT: Lee, Ming
; APPLICANT: Morrow, Jan E.
; APPLICANT: Welsh, Piri L.
; APPLICANT: Leon, Pedro E.
; TITLE OF INVENTION: Modulators of Actin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/333,735
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/080,897
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UW97-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1248 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-323-735-2

```

```

Query Match 13.3%; Score 136.5; DB 3; Length 1248;
Best Local Similarity 32.3%; Pred. No. 0.0029;
Matches 43; Conservative 8; Mismatches 59; Indels 23; Gaps 5;

QY 35 PLPQQ-----PPANQDNSSQTLQPTPIAPAPKPA-----APRDLRES----- 80
DB 578 PLPGSGTIIPPPAPGD-----STTPPPPPPPPPPPPLPGTAISPPPLSGDATIPPP 633

QY 81 -PGVENKLIPSVGSPASSTPLPDGTGPNSTPNRAVTVSGSNSSADPKAPPPPPVS 139
DB 634 PPLPEGVIGPSPLSGGTAIP-----PPLPGSARIPPPPPPLPGSAGIPIPPPPPLGE 689

QY 140 SGEPPTLGENPDG 152
DB 690 AGMPPTPPPLPGG 702

```

```

RESULT 13
US-08-818-112-53
; Sequence 53, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

```

```

; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-53

```

```

Query Match 13.1%; Score 135; DB 3; Length 332;
Best Local Similarity 27.5%; Pred. No. 0.00084;
Matches 42; Conservative 15; Mismatches 57; Indels 38; Gaps 6;

QY 6 HQINSNKTERSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQTLQPTPIAP 65
DB 7 HMQVDPNLTRKGRGLAALAIAAMAS-ASLTVAVPATANADP-----EPAPVPPT 57

QY 66 APKD-----AAPRDLRESFGVENKLIPSVGSPASSTPLPDGTGPNSTPNRAVTVSQ 121
DB 58 AASPSTAAAPPAP-----ATPVAPPPPPAAANTPN-----AQGGDP 93

QY 122 GSNSSADPKAPPPPPVSSGEP-PTLGENPDG 152
DB 94 NAAPPADPNAPPVPPVIAPNAPQVRIDNPVG 125

```

```

RESULT 14
US-08-818-111-53
; Sequence 53, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```


Result No.	Score	Query Match	Length	DB	ID	Description
1	1028	100.0	1426	15	US-10-322-579-15	Sequence 15, Appl
2	178	17.3	35	15	US-10-322-579-5	Sequence 5, Appl
3	167.5	16.3	707	11	US-09-919-039-278	Sequence 278, Appl
4	162.5	15.8	574	12	US-10-168-097A-76	Sequence 76, Appl
5	162.5	15.8	574	12	US-10-239-431A-38	Sequence 38, Appl
6	158	15.4	802	9	US-09-823-240-2	Sequence 2, Appl
7	154.5	15.0	731	14	US-10-086-464-17	Sequence 17, Appl
8	152	14.8	509	12	US-10-021-660-86	Sequence 86, Appl
9	151.5	14.7	503	14	US-10-078-547-2	Sequence 2, Appl
10	150.5	14.6	533	11	US-09-932-226-2	Sequence 2, Appl
11	150.5	14.6	533	15	US-10-098-184-2	Sequence 2, Appl
12	150	14.6	5179	9	US-09-932-317-1068	Sequence 1068, Ap
13	150	14.6	5179	10	US-09-833-263-1068	Sequence 1068, Ap
14	150	14.6	5179	14	US-10-025-380-1068	Sequence 1068, Ap
15	149	14.5	507	14	US-10-078-547-24	Sequence 24, Appl

QY 121 QGSSSSADKAPPPPPVSSGPPPTLGENPDGLSQEQLHRSRSLQTLRDQMLFPDEK 180
 |||||
 Db 319 QGSSSSADKAPPPPPVSSGPPPTLGENPDGLSQEQLHRSRSLQTLRDQMLFPDEK 378
 |||||
 QY 181 EFTGAQSGGPOQNP 194
 |||||
 Db 379 EFTGAQSGGPOQNP 392

RESULT 2
 US-10-322-579-5
 ; Sequence 5, Application US/10322579
 ; Publication No. US20030114413A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BASLER, Konrad
 ; APPLICANT: BRUNNER, Erich
 ; APPLICANT: FROESCH, Barbara
 ; APPLICANT: KRAMPS, Thomas
 ; APPLICANT: PETER, Oliver
 ; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY
 ; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON
 ; FILE REFERENCE: Q60361
 ; CURRENT APPLICATION NUMBER: US/10/322,579
 ; CURRENT FILING DATE: 2002-12-19
 ; PRIOR APPLICATION NUMBER: US/09/915,543
 ; PRIOR FILING DATE: 2001-07-27
 ; PRIOR APPLICATION NUMBER: 60/221,502
 ; PRIOR FILING DATE: 2000-07-28
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 5
 ; LENGTH: 35
 ; TYPE: PRT
 ; ORGANISM: Human lgs/bcl9
 US-10-322-579-5

Query Match 17.3%; Score 178; DB 15; Length 35;
 Best Local Similarity 100.0%; Pred. No. 2.4e-05;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 151 DGLSQEQLHRSRSLQTLRDQMLFPDEKFTGA 185
 |||||
 Db 1 DGLSQEQLHRSRSLQTLRDQMLFPDEKFTGA 35

RESULT 3
 US-09-919-039-278
 ; Sequence 278, Application US/09919039
 ; Publication No. US20030108871A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kaser, Matthew R.
 ; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
 ; FILE REFERENCE: PA-0035 US
 ; CURRENT APPLICATION NUMBER: US/09/919,039
 ; CURRENT FILING DATE: 2002-09-09
 ; PRIOR APPLICATION NUMBER: 60/222,113
 ; PRIOR FILING DATE: 2000-07-28
 ; NUMBER OF SEQ ID NOS: 401
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 278
 ; LENGTH: 707
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc:feature
 ; OTHER INFORMATION: Incyte ID No. US20030108871A1 2457215CD1
 US-09-919-039-278

Query Match 16.3%; Score 167.5; DB 11; Length 707;
 Best Local Similarity 40.0%; Pred. No. 0.0024;
 Matches 54; Conservative 5; Mismatches 47; Indels 29; Gaps 8;

QY 30 RNDPKLPQPPAPANQDN--SSQNRLOTPPIAPAPKP-----AAPRLDRESPGV 83
 :|||
 Db 53 QSGPKP-PIDPPPHQQQQPPPPQPPPHQPPPHQPPPPQDSSKPYV 111
 |||||
 QY 84 EN--KLIPSYGS--PASSTPLP-----PDGTGPNSTPN-NRAYTPVSGSNSSSADP 130
 |||||
 Db 112 AOGPGAPGVGSAPPASSAPPATPTTSGAPPGCGPCTPTPPPAVT-----SAPP 162
 |||||
 QY 131 KAPPPPPVSSGEPPT 145
 |||||
 Db 163 GAPPPTPPSSGVPT 177

RESULT 4
 US-10-168-097A-76
 ; Sequence 76, Application US/10168097A
 ; Publication No. US20030166245A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CENTRE NATIONAL OF THE RECHERCHE SCIENTIFIQUE
 ; APPLICANT: INSTITUT CURIE
 ; TITLE OF INVENTION: WASP FAMILY PROTEIN FRAGMENTS, AND USES THEREOF
 ; FILE REFERENCE: IFB99WASP
 ; CURRENT APPLICATION NUMBER: US/10/168,097A
 ; CURRENT FILING DATE: 2002-06-17
 ; NUMBER OF SEQ ID NOS: 84
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 76
 ; LENGTH: 574
 ; TYPE: PRT
 ; ORGANISM: Schizosaccharomyces pombe
 US-10-168-097A-76

Query Match 15.8%; Score 162.5; DB 12; Length 574;
 Best Local Similarity 30.9%; Pred. No. 0.0041;
 Matches 50; Conservative 21; Mismatches 64; Indels 27; Gaps 7;
 QY 7 IONISNNKTERSTA-----PLNTQISA---LRNDPKLPQPPAPANQDQ-----NSSQNT 54
 |:|||
 Db 275 IAPVSMNPAINSTKPLPPPSRVSAALAAANKRPPPPPPFRRNRKGPPIGNGSSNS 334
 |||||
 QY 55 RLQPTPP-----IPAPAPKPAAPRLDRESPGVENKLIPIVSGSPA--SSTPLPPDG 104
 |||||
 Db 335 SLPPPPPPPSNAAGSILPPQGRSAPPPPPPSRAGPTGRQPLSSRAVSNPPAPPA 394
 |||||
 QY 105 TGPSTNNRAYTPVSGSNSSADPKAPPPPPVSSGEPPTL 146
 |||||
 Db 395 IPGRSAP---ALPPLGNASRTST--PPVPTPPSLPPSAPPSL 431

RESULT 5
 US-10-239-431A-38
 ; Sequence 38, Application US/10239431A
 ; Publication No. US20030170726A1
 ; GENERAL INFORMATION:
 ; APPLICANT: FRADELIZE, JULIE
 ; APPLICANT: FRIEDERICH, EVELYNE
 ; APPLICANT: GOLSTEYN, ROY M.
 ; APPLICANT: LOUARD, DANIEL
 ; APPLICANT: NOIREAUX, VINCENT
 ; APPLICANT: SYKES, CECILE
 ; TITLE OF INVENTION: PEPTIDE SEQUENCES COMPRISING ONE OR MORE UNITS BINDING
 ; TITLE OF INVENTION: TO PROTEINS OF THE Ena/VASP FAMILY, AND THEIR USES
 ; FILE REFERENCE: 0508-1032
 ; CURRENT APPLICATION NUMBER: US/10/239,431A
 ; CURRENT FILING DATE: 2002-09-23
 ; PRIOR APPLICATION NUMBER: PCT/FR01/00843
 ; PRIOR FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: FR 00/03637
 ; PRIOR FILING DATE: 2000-03-22
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 38
 ; LENGTH: 574

[illegible]

Db 153 APPPPPPSRAGSGAGALPCAGHTVRRRTSSPRSSPPLSGPPGRASGRARPPPLLRAA 212
QY 110 -TPNNRAVTPVSGSNSSADKAPPPPPVSSGE-----PP-----TL 146
Db 213 PTPSPRALAP-----AAASPPPPPPPPGREGKKRKFPPGSSGSGTQTGAAAVAAAL 265
QY 147 GENPDGLSQEQLHRSRSLQIRDIQMLFDEKEFTGASG 188
Db 266 GSSPG-----RRLLPLLLRVGR-----PRSGAASG 291

RESULT 9

US-10-078-547-2
; Sequence 2, Application US/10078547
; Publication No. US20020199211A1
; GENERAL INFORMATION:
; APPLICANT: Narayanaswamy Ramesh
; APPLICANT: Miguel A. de la Fuente
; APPLICANT: Ines M. Anton
; APPLICANT: Raif S. Geba
; TITLE OF INVENTION: WIP, A WASP-Associated Protein
; FILE REFERENCE: 1242.1022-005
; CURRENT APPLICATION NUMBER: US/10/078,547
; PRIOR FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: 09/599,287
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: PCT/US98/27501
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/101,457
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/068,533
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Human
US-10-078-547-2

Query Match 14.7%; Score 151.5; DB 14; Length 503;
Best Local Similarity 26.5%; Pred. No. 0.019; Mismatches 84; Indels 57; Gaps 11;
Matches 60; Conservative 25;
QY 12 NNK-----TERSTAPLNTQISLRNDPKLPQQPPAPANQDQNS-----SONTRL-OPT 59
Db 289 NNKPPVPSTPRPSAPHRPH---LRPPPSRPGPPPLPPSSGNDPTPLPQNLSLSST 345
QY 60 PPIPAKPAPAPRPLDRESQVENKLIPSVGSPASSTPLPDGTGPNSTNNRAV--- 116
Db 346 PPLPSPGRSGPLPPPPPSRPPPPVRD-----PPGRSGPLPPPPVSRNGSTSRALPAT 398
QY 117 --TPVSGGSSSADPK-----APPPPPVSSGPPPTLGENPDGLSQEQLHR-- 161
Db 399 PQLPSRSGVDSRPSRPPPLPDPRSAGAPPPPPSTSRNGFQDSP---CEDEWESRY 455
QY 162 -----ERSLQIRDIQMLFDEKEFTGA----QSGGQQNP 194
Db 456 FHPISDLPPPEPVYQTKSVPSKLARNESR-SGSNRRRERGPPPLP 500

RESULT 10

US-09-922-226-2
; Sequence 2, Application US/09922226
; Publication No. US20030077664A1
; GENERAL INFORMATION:
; APPLICANT: Zhao, Yi
; APPLICANT: Thacher, Scott M.
; APPLICANT: Xiao, Jia-Hao
; APPLICANT: Kusari, Jyotirmoy
; APPLICANT: Chandraratna, Roshantha A.
; TITLE OF INVENTION: Methods of Screening For Compounds That
; TITLE OF INVENTION: Modulate Hormone Receptor Activity
Query Match 14.6%; Score 150.5; DB 15; Length 533;
Best Local Similarity 31.0%; Pred. No. 0.023; Mismatches 45; Indels 33; Gaps 5;
Matches 40; Conservative 11;
QY 31 NDPKPLPQ--QPPAPANQDQNSQNTLRLOTPPIPA-----PKPAAPRPLDRESP 81
Db 86 SSPNPLPQGVPPSPPG-----PPLPPTAPSLGSGGARGPPPPMPPPLGSPFP 134
QY 82 GVENKLIPSVGSPASSTPLPDGTGPNSTNNRAVTPVSGSNSSSADPKAP----- 133
Db 135 VISS-----SMGSPGLPPAPPFGSPVSSPQINSTVSLPGGSGPPEDVKPPVLGVRGLH 190
QY 134 -PPPPVSSG 141

FILE REFERENCE: P-AR 4681
; CURRENT APPLICATION NUMBER: US/09/922,226
; CURRENT FILING DATE: 2002-01-09
; PRIOR APPLICATION NUMBER: US 60/284,797
; PRIOR FILING DATE: 2001-04-18
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-226-2

Query Match 14.6%; Score 150.5; DB 11; Length 533;
Best Local Similarity 31.0%; Pred. No. 0.023; Mismatches 45; Indels 33; Gaps 5;
Matches 40; Conservative 11;
QY 31 NDPKPLPQ--QPPAPANQDQNSQNTLRLOTPPIPA-----PKPAAPRPLDRESP 81
Db 86 SSPNPLPQGVPPSPPG-----PPLPPTAPSLGSGGARGPPPPMPPPLGSPFP 134
QY 82 GVENKLIPSVGSPASSTPLPDGTGPNSTNNRAVTPVSGSNSSSADPKAP----- 133
Db 135 VISS-----SMGSPGLPPAPPFGSPVSSPQINSTVSLPGGSGPPEDVKPPVLGVRGLH 190
QY 134 -PPPPVSSG 141
Db 191 CPEPPGGPG 199

RESULT 11

US-10-098-184-2
; Sequence 2, Application US/10098184
; Publication No. US20030105333A1
; GENERAL INFORMATION:
; APPLICANT: Pfahl, Magnus
; APPLICANT: Tachdjian, Catherine
; APPLICANT: Al-Shamma, Hussien A.
; APPLICANT: Fanjul, Andrea
; APPLICANT: Pleyne, David P.M.
; APPLICANT: Spruce, Lyle W.
; APPLICANT: Fine, Richard
; APPLICANT: Zapf, James W.
; TITLE OF INVENTION: RXR ACTIVATING MOLECULES
; FILE REFERENCE: 13099.0016U2
; CURRENT APPLICATION NUMBER: US/10/098,184
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: 60/274,342
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. US20030105333A1e -
; OTHER INFORMATION: synthetic construct
US-10-098-184-2

Query Match 14.6%; Score 150.5; DB 15; Length 533;
Best Local Similarity 31.0%; Pred. No. 0.023; Mismatches 45; Indels 33; Gaps 5;
Matches 40; Conservative 11;
QY 31 NDPKPLPQ--QPPAPANQDQNSQNTLRLOTPPIPA-----PKPAAPRPLDRESP 81
Db 86 SSPNPLPQGVPPSPPG-----PPLPPTAPSLGSGGARGPPPPMPPPLGSPFP 134
QY 82 GVENKLIPSVGSPASSTPLPDGTGPNSTNNRAVTPVSGSNSSSADPKAP----- 133
Db 135 VISS-----SMGSPGLPPAPPFGSPVSSPQINSTVSLPGGSGPPEDVKPPVLGVRGLH 190
QY 134 -PPPPVSSG 141

```
Db 191 CPPPGGPG 199
|||||
Query Match 14.6%; Score 150; DB 10; Length 5179;
Best Local Similarity 25.6%; Pred. No. 0.24;
Matches 41; Conservative 20; Mismatches 65; Indels 34; Gaps 5;

RESULT 12
US-09-922-217-1068
; Sequence 1068, Application US/09922217
; Patent No. US2002007641A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1068

Query Match 14.6%; Score 150; DB 9; Length 5179;
Best Local Similarity 25.6%; Pred. No. 0.24;
Matches 41; Conservative 20; Mismatches 65; Indels 34; Gaps 5;

QY 15 TERSTAPLNTQISALRNDPKPLPQQPPAPANDQSSQNTRLQPTPIAPAPKPAAP-- 72
Db 1428 TTTTPPTTTPSPPIITTTTTPPTTTPSPPISTTTTTPPT-TTPSPPTTTPSPPTTTPSP 1486
QY 73 -----PRPLDRESGVENKLIPSGSPASSTPLPPDGT-----GPNSTPNNR 114
Db 1487 PTTTTPPTTTPSPMTTPTP-----PASTTTLPTTTPSPPTTTPPTTTPSPPT 1542
QY 115 AVTPVSGQSSSADPKAPP-----PPPVSSGEPT 145
Db 1543 TTTPTTPTSTTTLPTTTPSPPTTTPPTTTPPTTTPSPPT 1582
QY 115 AVTPVSGQSSSADPKAPP-----PPPVSSGEPT 145
Db 1543 TTTPTTPTSTTTLPTTTPSPPTTTPPTTTPPTTTPSPPT 1582

RESULT 13
US-09-833-263-1068
; Sequence 1068, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.
; APPLICANT: Meagher, Madeleine J.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833,263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-263-1068

Query Match 14.6%; Score 150; DB 14; Length 5179;
Best Local Similarity 25.6%; Pred. No. 0.24;
Matches 41; Conservative 20; Mismatches 65; Indels 34; Gaps 5;

QY 15 TERSTAPLNTQISALRNDPKPLPQQPPAPANDQSSQNTRLQPTPIAPAPKPAAP-- 72
Db 1428 TTTTPPTTTPSPPIITTTTTPPTTTPSPPISTTTTTPPT-TTPSPPTTTPSPPTTTPSP 1486
QY 73 -----PRPLDRESGVENKLIPSGSPASSTPLPPDGT-----GPNSTPNNR 114
Db 1487 PTTTTPPTTTPSPMTTPTP-----PASTTTLPTTTPSPPTTTPPTTTPPTTTPSPPT 1542
QY 115 AVTPVSGQSSSADPKAPP-----PPPVSSGEPT 145
Db 1543 TTTPTTPTSTTTLPTTTPSPPTTTPPTTTPPTTTPSPPT 1582

RESULT 14
US-10-025-380-1068
; Sequence 1068, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Derrick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C14
; CURRENT APPLICATION NUMBER: US/10/025,380
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-025-380-1068

Query Match 14.6%; Score 150; DB 14; Length 5179;
Best Local Similarity 25.6%; Pred. No. 0.24;
Matches 41; Conservative 20; Mismatches 65; Indels 34; Gaps 5;

QY 15 TERSTAPLNTQISALRNDPKPLPQQPPAPANDQSSQNTRLQPTPIAPAPKPAAP-- 72
Db 1428 TTTTPPTTTPSPPIITTTTTPPTTTPSPPISTTTTTPPT-TTPSPPTTTPSPPTTTPSP 1486
QY 73 -----PRPLDRESGVENKLIPSGSPASSTPLPPDGT-----GPNSTPNNR 114
Db 1487 PTTTTPPTTTPSPMTTPTP-----PASTTTLPTTTPSPPTTTPPTTTPPTTTPSPPT 1542
QY 115 AVTPVSGQSSSADPKAPP-----PPPVSSGEPT 145
Db 1543 TTTPTTPTSTTTLPTTTPSPPTTTPPTTTPPTTTPSPPT 1582

RESULT 15
US-10-078-547-24
; Sequence 24, Application US/10078547
; Publication No. US20020199211A1
; GENERAL INFORMATION:
```

Search completed: October 15, 2003, 10:50:59
Job time : 83.2802 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	177.5	17.3	473	2	S50755		hypothetical prote
2	168.5	16.4	876	2	T45801		hypothetical prote
3	167.5	16.3	707	2	A43302		PTB-associated spl
4	162.5	15.8	574	2	T43556		Wiskott-Aldrich sy
5	162.5	15.8	574	2	T38819		wiskott-aldrich sy
6	162.5	15.8	1151	2	T18535		high molecular mas
7	161.5	15.7	815	2	B56708		extracellular sign
8	160	15.6	839	2	F75518		hypothetical prote
9	158.5	15.4	351	2	S50754		hypothetical prote
10	158.5	15.4	1188	2	S49915		extensin-like prot
11	157.5	15.3	464	2	S22697		extensin - Volvox
12	156	15.2	449	2	S16748		proline-rich prote
13	154.5	15.0	731	2	T04455		proline-rich prote
14	152.5	14.8	379	2	T05441		proline-rich prote
15	152.5	14.8	379	2	D85257		extensin-like prot
16	150.5	14.6	251	1	PHUFP		salivary proline-r
17	150.5	14.6	533	2	S37781		retinoid X recepto
18	150	14.6	3020	2	A43932		mucin 2 precursor,
19	149	14.5	188	2	D29149		proline-rich prote
20	149	14.5	302	2	T32711		hypothetical prote
21	149	14.5	426	2	J01696		pistil extensin-li
22	149	14.5	1201	2	G86441		unknown protein [i
23	148.5	14.4	403	2	S2796		prp12 protein - hu
24	148.5	14.4	731	2	B86369		hypothetical prote
25	147	14.3	348	1	OZQBK		circumsporozoite p
26	147	14.3	416	2	T34279		hypothetical prote
27	146.5	14.3	393	2	PQ0479		pistil extensin-li
28	146.5	14.3	708	2	N96711		hypothetical prote
29	146	14.2	740	2	B24264		proline-rich prote

A;Gene: NCSP:B11B22.30
A;Map position: 6
A;Introns: 75/3; 190/1

RESULT 4
T43556
Wiskott-Aldrich syndrome protein homolog - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C;Accession: T43556
R;Zankel, T.C.; Ow, D.W.
submitted to the EMBL Data Library, December 1997
A;Description: A Wiskott-Aldrich Syndrome protein homolog in Schizosaccharomyces pombe
A;Reference number: Z22575
A;Accession: T43556
A;Status: preliminary; translated from GB/EMBL/DDBB
A;Molecule type: DNA
A;Residues: 1-574 <FAN>
A;Cross-references: EMBL:AF038575; PIDN:AAB92587.1
A;Experimental source: strain JS21
C;Genetics:
A;Gene: wsp1
A;Map position: I
A;Introns: 72/3; 519/3; 564/1

```
Query Match      15.8%; Score 162.5; DB 2; Length 574;  
Best local Similarity 30.9%; Pred. No. 0.017;  
Matches 50; Conservative 21; Mismatches 64; Indels 27; Gaps 7;  
  
QY    7 IQNISNKKTERSTA-----PLNTQISA---LRNDKPLPQQPPAPANQDQ-----NSSQNT 54  
       ||| ||| | : : : : | : | : | : | : | : | : | : | : | : | :  
Db     275 IAVSMNPAINSTSKPLPPLPPSSSRVSAALAAANKKRPPPPPPPSRRNRGPPITGGSSNS 334  
       ||| ||| | : : : : | : | : | : | : | : | : | : | : | : | :  
  
QY    55 RLQPTTP-----IPAPAKPAAPPRLDRSPGVENKLIPSGVSPA-STTPIPPDG 104  
       ||| ||| | : : : : | : | : | : | : | : | : | : | : | : | :  
Db     335 SLRPPPPPPSRNAAGSIPLDPQGRASAPPPPPPPPSAISTGTGRPPLSSRAVSNNPAPPPA 394
```

```
QY      105 TGNSTPNNAVTPVCGSSNSSADPKAPPPPPVSSEGPDTL 146
       |   | :   | :   | :   | :   | :   | :   |||
Db      395 IPRGSAPE---ALPPLGNASTST--PPVPTPSLPSPASPSL 431
```

RESULT 5
T38819

wiskott-aldrich syndrome protein homolog 1 - fission yeast (*Schizosaccharomyces pombe*)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T38819
R;Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z21813
A;Accession: T38819

A;Accession: J06019
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-574 <CON>
A;Cross-references: EMBL:Z98980; NID:el060691; PIDN:CAB11718.1; GSPDB:GN00066; SPDB:SP000006
A;Experimental source: strain 972h-; cosmid c4F10
C:Genetics:
A;Gene: wsp1; SPDB:SPAC4F10.15c

```
A:Map position: 1
A:Introns: 72/3; 519/3; 564/1

Query Match      15.8%; Score 162.5; DB 2; Length 574;
Best Local Similarity 30.3%; Pred. NO. 0.017;
Matches 50; Conservative 21; Mismatches 64; Indels 27; Gaps 7;

QY    7 IONISNKKTERSTA-----PLATQISA---LRNDPKFLPQQPPAPANODQ----NSSONT 54
      | : | ||| : : : ||| : | : ||| : | : ||| : | : ||| : | : ||| :
Db     275 IAPVSMNPAINSTSKPLPPSPSVKSAALAANKKRPPPPPPSRNRGRKPPIGNSSNS 334

QY    55 RLQPTTP-----IPAPAKPAAPPALDREESPGVENKLIPSVGSPA-SSTPLPPDG 104
      | : | ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| :
Db     335 SUPPPPPPPPSRNAAGSILPQPGRSAPPPPPPPPSAFSTGQRPPPLLSSRAVSNNPPAPPPA 394

QY    105 TGTNSTPNRNVTPVQSGNSSSADPKAPPPPPVSSGEPPTL 146

Db     395 IFQRSAP--ALPPLGNASRTST--PPVPTPFSLPSPAPSLSL 431
```

RESULT 6
T18535
high molecular mass nuclear antigen - chicken (fragment)
C:Species: Gallus gallus (chicken)
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18535
R:Shimada, K.; Harata, M.; Mizuno, S.
J. Cell Sci. 110, 3031-3041, 1997
A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chick
A:Reference number: Z18955; MUID:9803440; PMID:9365273
A:Accession: T18535
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1151 <SH2>
A:Cross-references: EMBL:D88440; NID:d1171138; PID:d1025045; PIDN:BA24137.1

Query Match 15.8%; Score 162.5; DB 2; Length 1151;
Best Local Similarity 33.1%; Pred. No. 0.033;
Matches 45; Conservative 12; Mismatches 40; Indels 39; Gaps 6;
QY 33 PKPLQQPPAPANQONSSQNTLRQTP-PIFAPAPKPAAPRP-----IDRSPGVGE 84
Db 3 PTPPPRNPPTP-----PPAPSPAPAPAPTAPPRPKWVPIAELHPAAPQPP 49
QY 85 NKLIPSVGSPASSTPLPDGTGPNSTNNRATVPVSGNSSSADPKAP---PPPVSSG 141
Db 50 PKWVPIGAP-----PPPGTEP-----TPFSKPTDGADAPKASAEITSPPPASP 94
QY 142 EPPTLGENPDGLSQEQ 157
Db 95 SPFDGPKAPSGAGEAE 110

RESULT 7
B56708
extracellular signal-regulated kinase 5 - human
C:Species: Homo sapiens (man)
C>Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 24-Sep-1999
C:Accession: B56708
R:Zhou, G.; Bao, Z.Q.; Dixon, J.E.
J. Biol. Chem. 270, 12665-12669, 1995
A:Title: Components of a new human protein kinase signal transduction pathway.
A:Reference number: A56708; MUID:95279403; PMID:7759517
A:Accession: B56708
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-815 <ZHO>
A:Cross-references: GB:U25278; NID:g837260; PIDN:AAA81381.1; PID:g837261
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
F:52-315/Domain: protein kinase homology <KIN>
F:60-68/Region: protein kinase ATP-binding motif

Query Match 15.7%; Score 161.5; DB 2; Length 815;
Best Local Similarity 32.1%; Pred. No. 0.027;
Matches 54; Conservative 7; Mismatches 64; Indels 43; Gaps 7;
QY 17 RSTAPLNTQISALRNDKPLP-----QQPPAPANQONSSQNTLRQTPPIP 63
Db 576 RPAAPALSVAPAPAPTPTTPVQPTSPPGPLAQPTGPOPSAGSISGVPVACPPP 635
QY 64 APAPKPAAPPLDRESFG-----VENKLIPSVGSPASSTP-----LPPDGT 105
Db 636 GPAPHTGPPGPIPVAPAPQIATSTSLAAQSLVPPGLPGSSTGVLFYPPGLPPDDA 695
QY 106 G--PNS-----TPNNRATVPVSGNSSSADPKAPPPPVSSGEPPTLG 147
Db 696 GGAPQSSMSSEFDNVLVT--QQLSKSQVEDFL-----PPVSGTPKGSG 737

RESULT 8

F75518
hypothetical protein - Deinococcus radiodurans (strain RL)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: F75518
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans RL.
A:Reference number: A75250; MUID:2003696; PMID:10567266
A:Accession: F75518
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-839 <WH1>
A:Cross-references: GB:AE001904; GB:AE000513; NID:g6458129; PIDN:AAF10038.1; PID:g64
A:Experimental source: strain RL
C:Genetics:
A:Gene: DR0458
A:Map position: 1
C:Superfamily: herpesvirus immediate-early protein IE110; RING finger homology

Query Match 15.6%; Score 160; DB 2; Length 839;
Best Local Similarity 31.3%; Pred. No. 0.033;
Matches 52; Conservative 13; Mismatches 79; Indels 22; Gaps 7;
QY 33 PKP-LPQQPPAPANQONSSQNTLRQ-----TPPIFAPA---PKPAAP-PRPLDRESGVGE 84
Db 222 PTFNPAQTTPAPATAPAAQAQTPTAQAPATQTATPAAQAQRPAAGGAPSPAPAPAQANAPA 281
QY 85 NKLIPSVGSPASSTPLPDGTGPNSTNNRATVPVSGS-NSSSADPKAPPPPVSSGHP 143
Db 282 GSVPEAVPESSTPAFSAQTPTPTRETATGTAESPAPNSAAPNEPASEPV-AGRP 340
QY 144 PTLGENPDGLSQEQLEHRSLSQTLRLDIQRMFLPDPEKEFTGAQSGG 189
Db 341 GTAASSPESASPTVTTPRGET-----PDTAASAGTFSAG 374

RESULT 9
S50754
hypothetical protein WP6 - Chlamydomonas eugametos
C:Species: Chlamydomonas eugametos
C>Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000
C:Accession: S50754
R:Woessner, J.P.; Molendijk, A.J.; van Egmond, P.; Klis, F.M.; Goodenough, U.W.; Har
Plant Mol. Biol. 26, 947-960, 1994
A:Title: Domain conservation in several volvocalean cell wall proteins.
A:Reference number: S50754; MUID:95093034; PMID:8000007
A:Accession: S50754
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-351 <WOE>
A:Cross-references: EMBL:L29028; NID:g530877; PIDN:AAB53954.1; PID:g530878

Query Match 15.4%; Score 158.5; DB 2; Length 351;
Best Local Similarity 30.9%; Pred. No. 0.017;
Matches 46; Conservative 15; Mismatches 71; Indels 17; Gaps 5;
QY 9 NISNNKTRSTAPLNTQ-----ISALRNDKPLPQQPPAPANQONS-----SQNTLRQPT 59
Db 144 SYVNNCIDTRAPYNCSTFNTSVITPTPSPSPSPSPSPSPSPSPSPSPSKASPS 203
QY 60 P-PIFAPAPKPAAPPRPLDRESGVENKLIPSVGSPASSTPLPPDGTGPNSTNNRATVP 118
Db 204 PSPKASPSPSKASAPSPQSPPTSPKASPVASVQSQPTSPRSPRSPPTSP-----SP 258
QY 119 VSQGSNSSSADPKAP---PPPVSSGEPPT 145
Db 259 SPKASPPPSASPSASPSLSPKVSPSTPPT 287

RESULT 10

S49915
extensin-like protein - maize
C:Species: Zea mays (maize)
C>Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
C:Accession: S49915
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A:Description: Pex genes: pollen-specific genes with extensin-like domains.
A:Reference number: S49915
A:Accession: S49915
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1188 <RUB>
A:Cross-references: EMBL:Z34465; NID:g600117; PIDN:CAA84230.1; PID:g600118

Query Match 15.4%; Score 158.5; DB 2; Length 1188;
Best Local Similarity 32.0%; Pred. No. 0.057;
Matches 49; Conservative 17; Mismatches 68; Indels 19; Gaps 7;

Qy 14 KTERSTAPLNQTISALNDPKPLP--QQPPAPANQDNSSQNTRLOP---TPPIPAFAPK 68
Db ||| |::| | | | | | | | | | | | | | | | | | | |
961 KSGPPPPAVNLPPPEVKSSPPPTPVSPPPAPKSSPPPADMSSPPPPPEVKSPPPPAPVSS 1020
Qy 69 P-----AAPRLDRESPOVENKLIPSVGSPPASSFTLP---PDGIGNSTPNNAVTVPV 119
Db ||| | | | | | | | | | | | | | | | | | | | | |
1021 PPPPVKSPPPPAVSSPPPPVKSS---PPPPAVSSPPPPPVKSPPPPAPITSSPPPPVKSP 1077
Qy 120 SQGSNSSADDP-KAPPDP-PVSSGEPTLTGENP 150
Db || | | | | | | | | | | | | | | | | | | | | | |
1078 PPAPVSSPPPPVKSPPPPPAPVSSPPPPIKSPPP 1110

RESULT 11
S22697
extensin - Volvox carteri (fragment)
C:Species: Volvox carteri
C>Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 21-Jul-2000
C:Accession: S22697; S21006
R:Ertl, H.; Hallmann, A.; Wenzl, S.; Sumper, M.
EMBO J. 11, 2055-2062, 1992
A>Title: A novel extensin that may organize extracellular matrix biogenesis in Volvox ca
A:Reference number: S22697; PMID:92289669; PMID:1600938
A:Accession: S22697
A:Molecule type: mRNA
A:Residues: 1-464 <HAL>
A:Cross-references: EMBL:X65165; NID:g21991; PIDN:CAA46283.1; PID:g21992
C:Keywords: glycoprotein

Query Match 15.3%; Score 157.5; DB 2; Length 464;
Best Local Similarity 31.3%; Pred. No. 0.026;
Matches 41; Conservative 11; Mismatches 64; Indels 15; Gaps 3;

Qy 21 PLNTQISALNRDPKPLPQQPPAPANQDNSSQNTRLOTPPIPAPAKPAAPRPDLDRS 80
Db | : : | | | | | | | | | | | | | | | | | | | | |
292 PPPPRVSPSPPPQPVSSPPPPPPPPPSPPSPSPPPPPSPSPSPPPPPPPSPSPSP 351
Qy 81 PGVENKLIPSVGSFASTPLPDGTGNSTPNNAVTVPVSOGSNSSADDPKAPPP-PPVS 139
Db | | | | | | | | | | | | | | | | | | | | | | |
352 P---PRSSPSPPPVVSPPPPPPPRASPPPPP-----ASSPPPPRRPPPPSPPS 397
Qy 140 SGEPPTLTGENP 150
Db | | | | | | | | | | | | | | | | | | | | | | |
398 PPPPATATAANP 408

RESULT 12
S16748
proline-rich protein - rape (fragment)
C:Species: Brassica napus (rape)
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Sep-1997
C:Accession: S16748
R:Robertis, M.R.; Foster, G.D.; Blundell, R.P.; Robinson, S.W.; Draper, J.; Scott, R.J.
submitted to the EMBL Data Library, August 1991

A:Note: F7K2.50
C:Superfamily: hydroxyproline-rich glycoprotein

Query Match 14.8%; Score 152.5; DB 2; Length 379;
Best Local Similarity 31.8%; Pred. No. 0.041;
Matches 41; Conservative 12; Mismatches 45; Indels 31; Gaps 6;

QY 33 PKPLPQQPPAPANQDNSSONTLQTPPT---PAPAPKPAAPRPDLDRSGVENKLLIP 89
Db 51 PQDPQPPTPT-----FQAPFANDQPPPPQSTSP-PVATTPPALPPKPLP 98
QY 90 SVGSPASSTFLPDGTGPNSTPNNRVTPVSGNSSSADPKRAPPVPVSSGEPPTLGEN 149
Db 99 PPLSPQTTPPPPAITPPPP---AITP-----PLSPPPPAIT---PPPLPATT 142

QY 150 PDGLSQEQL 158
Db 143 PPALPPKPL 151

RESULT 15
D85257
extensin-like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-May-2002
C:Accession: D85257
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488; PMID:10617198
A:Accession: D85257
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-379 <STO>
A:Cross-references: GB:NC_001268; NID:g7269093; PIDN:CAB79202.1; GSPDB:GN00140
C:Genetics:
A:Gene: At4g22470
A:Map position: 4
C:Superfamily: hydroxyproline-rich glycoprotein

Query Match 14.8%; Score 152.5; DB 2; Length 379;
Best Local Similarity 31.8%; Pred. No. 0.041;
Matches 41; Conservative 12; Mismatches 45; Indels 31; Gaps 6;

QY 33 PKPLPQQPPAPANQDNSSONTLQTPPT---PAPAPKPAAPRPDLDRSGVENKLLIP 89
Db 51 PQDPQPPTPT-----FQAPFANDQPPPPQSTSP-PVATTPPALPPKPLP 98
QY 90 SVGSPASSTFLPDGTGPNSTPNNRVTPVSGNSSSADPKRAPPVPVSSGEPPTLGEN 149
Db 99 PPLSPQTTPPPPAITPPPP---AITP-----PLSPPPPAIT---PPPLPATT 142

QY 150 PDGLSQEQL 158
Db 143 PPALPPKPL 151

Job time : 48.0467 secs
Job completed: October 15, 2003, 10:33:12

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 23.4008 Seconds
(without alignments)
389.867 Million cell updates/sec

Title: US-09-915-543-15_COPY_199_392

Perfect score: 1028

Sequence: 1 TIVSFHIONISNNKTERSTA.....LFPDEKETGAQSGPQNP 194

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1024	99.6	1426	1	BCL9_HUMAN
2	167.5	16.3	707	1	SFQO_HUMAN
3	161.5	15.7	815	1	Q13164 homo sapien
4	159.5	15.5	555	1	Q03173 mus musculus
5	158	15.4	802	1	Q03173 mus musculus
6	156	15.2	449	1	Q03173 mus musculus
7	151.5	14.7	503	1	Q03173 mus musculus
8	150.5	14.6	251	1	Q03173 mus musculus
9	150.5	14.6	533	1	Q03173 mus musculus
10	150	14.6	5179	1	Q03173 mus musculus
11	149	14.5	426	1	Q03173 mus musculus
12	147	14.3	347	1	Q03173 mus musculus
13	146	14.2	534	1	Q03173 mus musculus
14	146	14.2	2167	1	Q03173 mus musculus
15	144.5	14.1	339	1	Q03173 mus musculus
16	144.5	14.1	542	1	Q03173 mus musculus
17	144	14.0	535	1	Q03173 mus musculus
18	144	14.0	817	1	Q03173 mus musculus
19	143.5	14.0	520	1	Q03173 mus musculus
20	143	13.9	296	1	Q03173 mus musculus
21	143	13.9	1206	1	Q03173 mus musculus
22	143	13.9	1468	1	Q03173 mus musculus
23	142.5	13.9	276	1	Q03173 mus musculus
24	141.5	13.8	283	1	Q03173 mus musculus
25	141.5	13.8	505	1	Q03173 mus musculus
26	141.5	13.8	620	1	Q03173 mus musculus
27	140.5	13.7	582	1	Q03173 mus musculus
28	140	13.6	806	1	Q03173 mus musculus
29	139.5	13.6	268	1	Q03173 mus musculus
30	139.5	13.6	505	1	Q03173 mus musculus
31	139.5	13.6	1664	1	Q03173 mus musculus
32	139	13.5	501	1	Q03173 mus musculus
33	138.5	13.5	2161	1	Q03173 mus musculus

34	138.5	13.5	3164	1	TEGU_HSV11
35	137.5	13.4	234	1	PRP2_MOUSE
36	137	13.3	261	1	PRP2_MOUSE
37	137	13.3	279	1	Y091_NPVOP
38	137	13.3	512	1	GAG_SMSAV
39	137	13.3	516	1	Y369_TREPA
40	136.5	13.3	458	1	RXRB_RAT
41	136.5	13.3	1248	1	DIAL_HUMAN
42	136.5	13.3	1409	1	AEX3_CAEEL
43	136.5	13.3	2142	1	BAT2_HUMAN
44	136	13.2	633	1	LAI1_YEAST
45	136	13.2	670	1	SYN1_MOUSE

ALIGNMENTS

RESULT 1

BCL9_HUMAN					
ID	BCL9_HUMAN	STANDARD;	PRT;	1426	AA.
AC	Q00512;				
DT	28-FEB-2003	(Rel. 41, Created)			
DT	28-FEB-2003	(Rel. 41, Last sequence update)			
DT	15-SEP-2003	(Rel. 42, Last annotation update)			
DE	B-cell lymphoma 9 protein (Bcl-9) (Legless homolog).				
GN	BCL9.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Fetal brain;				
RX	MEDLINE=98158621; PubMed=9490669;				
RA	Willis T.G., Zalcberg I.R., Coignet L.J.A., Wlodarska I., Stul M.,				
RA	Jadav D.M., Bastard C., Treleaven J.G., Catovsky D., Silva M.L.M.,				
RA	Dyer M.J.S.;				
RT	"Molecular cloning of translocation t(1;14)(q21;q32) defines a novel				
RT	gene (BCL9) at chromosome 1q21.";				
RL	Blood 91:1873-1881(1998).				
RN	[2]				
RP	FUNCTION.				
RC	MEDLINE=21952490; PubMed=11955446;				
RA	Kramps T., Peter O., Brunner E., Neilen D., Froesch B., Chatterjee S.,				
RA	Murone M., Zuelig S., Basler K.;				
RT	"Wnt/Wingless signaling requires BCL9/legless-mediated recruitment of				
RT	pygopus to the nuclear beta-catenin-TCF complex.";				
RL	Cell 109:47-60(2002).				
CC	-!- FUNCTION: Involved in signal transduction through the wnt pathway.				
CC	-!- SUBUNIT: Binds to beta-catenin (CTNBL1), PYGO1 and PYGO2.				
CC	-!- SUBCELLULAR LOCATION: Nuclear (Probable).				
CC	-!- TISSUE SPECIFICITY: Detected at low levels in thymus, prostate,				
CC	testis, ovary and small intestine, and at lower levels in spleen,				
CC	colon and blood.				
CC	-!- DISEASE: Involved in a t(1;14)(q21;q32) chromosomal translocation				
CC	found in a patient with precursor B-cell acute lymphoblastic				
CC	leukemia (ALL). This translocation leaves the coding region				
CC	intact, but may have pathogenic effects due to alterations in the				
CC	expression level of BCL9. Several cases of translocations within				
CC	the 3' untranslated region of BCL9 have been found in B-cell				
CC	malignancies.				
CC	-!- CAUTION: It is uncertain whether Met-1 or Met-27 is the initiator.				
CC	-!- CAUTION: Ref.1 sequence differs from that shown due to a				
CC	frameshift in position 1391.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				

QY 84 EN-KLIPSVGS--PASSTPLP-----PDGTGNSTPN-NRAVTPVSGSNSSADP 130
 Db 112 AOGPGAPGVGSAAPASSAPATPTSGAPGSGPTPTPPAVT-----SAPP 162
 QY 131 KAPPPPPVSSGEPTT 145
 Db 163 GAPPPTPPSSGVFTT 177

RESULT 3
 MK07_HUMAN
 ID MK07_HUMAN STANDARD; PRT; 815 AA.
 AC Q13164; Q16634;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Mitogen-activated protein kinase 7 (EC 2.7.1.-) (Extracellular signal-
 regulated kinase 5) (ERK-5) (ERK4) (BMK1 kinase).
 GN MAPK7 OR PRKM7 OR ERK5 OR ERK4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=95279403; PubMed=7759517;
 RA Zhou G., Bao Z.Q., Dixon J.E.;
 RT "Components of a new human protein kinase signal transduction
 pathway".
 RL J. Biol. Chem. 270:12665-12669(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=95374539; PubMed=7646528;
 RA Lee J.-D., Ulevitch R.J., Han J.;
 RT "Primary structure of BMK1: a new mammalian map kinase".;
 RL Biochem. Biophys. Res. Commun. 213:715-724(1995).
 CC -!- FUNCTION: MEK5 AND ERK5 INTERACT SPECIFICALLY WITH ONE ANOTHER AND
 NOT WITH MEK1/ERK1 OR MEK2/ERK2 PATHWAYS.
 CC -!- ENZYME REGULATION: Activated by tyrosine and threonine
 phosphorylation (By similarity).
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN MANY ADULT TISSUES. ABUNDANT IN
 HEART, PLACENTA, LUNG, KIDNEY AND SKELETAL MUSCLE. NOT DETECTABLE
 IN LIVER.
 CC -!- DOMAIN: THE SECOND PROLINE-RICH REGION MAY INTERACT WITH ACTIN
 TARGETING THE KINASE TO A SPECIFIC LOCATION IN THE CELL.
 CC -!- PTM: AUTOPHOSPHORYLATED ON THREONINE AND TYROSINE RESIDUES, WHEN
 THE C-TERMINAL PART OF THE KINASE, WHICH COULD HAVE A REGULATORY
 ROLE, IS ABSENT.
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC MAP KINASE SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U25278; AAA81381.1; -;
 DR EMBL; U29725; AAA82931.1; -;
 DR EMBL; U29726; AAA82932.1; -;
 DR EMBL; U29727; AAA82933.1; -;
 DR PIR; B56708; B56708.
 DR HSSP; P24941; IHCL.
 DR Genew; HGNC:6880; MAPK7.
 DR MIN; 602521; -;
 DR GO; GO:0004707; F:MAP kinase activity; TAS.
 DR GO; GO:0007165; P:signal transduction; TAS.
 DR InterPro; IPR003527; MAP_kin.

DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR002290; Ser_thr_kinase.
 DR Pfam; pf00069; pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00220; S_TKC; 1.
 DR PROSITE; PS01351; MAPK; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW Transferase; Serine/threonine-protein kinase; ATP-binding; Cell cycle;
 KW Phosphorylation.
 FT DOMAIN 54 346 PROTEIN KINASE.
 FT DOMAIN 337 340 POLY-ALA.
 FT DOMAIN 433 464 PRO-RICH 1.
 FT DOMAIN 520 523 POLY-ARG.
 FT DOMAIN 577 699 PRO-RICH 2.
 FT NP_BIND 60 68 ATP (BY SIMILARITY).
 FT BINDING 83 83 ATP (BY SIMILARITY).
 FT ACT_SITE 181 181 BY SIMILARITY.
 FT MOD_RES 218 218 PHOSPHORYLATION (ACTIVATES THE KINASE)
 FT MOD_RES 220 220 (BY SIMILARITY).
 FT MOD_RES 220 220 PHOSPHORYLATION (ACTIVATES THE KINASE)
 FT CONFLICT 19 32 (BY SIMILARITY).
 FT CONFLICT 609 609 AREGTRPRLCLCS -> GPVKVEPAHTAASVA (IN
 REF. 2).
 FT CONFLICT L -> V (IN REF. 2).
 SQ SEQUENCE 815 AA; 88636 MW; 379AD69803207CCF CRC64;
 Query Match 15.7%; Score 161.5; DB 1; Length 815;
 Best Local Similarity 32.1%; Pred. No. 0.069;
 Matches 54; Conservative 7; Mismatches 64; Indels 43; Gaps 7;

QY 17 RSTAPLNTQISALRNDKPLP-----QQPPAPANDQNSQNTRLQPTPIP 63
 Db 576 RPAAPALTSVPAPAPAPTPTPTVQTPSPGPGPLAQTPGQPSAGSTGVPVQACPPP 635
 QY 64 APAPKPAAPRPDLRESPG-----VENKLIPSGVSPASSTP-----LPPDGT 105
 Db 636 GPAPHTGPGPIPVAPPQTATSTSLAASLVPPGLPGSGSTGVLVPPFPLPPDDA 695
 QY 106 G--PNS-----TPNNRAVTPVSGSNSSADPKAPPPPVSSGEPPTLG 147
 Db 696 GGAQSSMSFSDVNLVT--QQLSKSQVEDPL----PPVFSSTPKSG 737

RESULT 4
 GPI_CHLRE
 ID GPI_CHLRE STANDARD; PRT; 555 AA.
 AC Q9FPQ6; Q03927;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Vegetative cell wall protein gpl precursor (Hydroxyproline-rich
 glycoprotein 1).
 DE glycoprotein 1).
 GN GPI.
 OS Chlamydomonas reinhardtii.
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Chlamydomonadaceae; Chlamydomonas.
 CC NCB1_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21159092; PubMed=11258910;
 RA Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,
 RA Goodenough U.W.;
 RT "Glycosylated polyproline II rods-with-kinks as a structural motif in
 plant hydroxyproline-rich glycoproteins".;
 RL Biochemistry 40:2978-2987(2001).
 RN [2]
 RP PARTIAL PRELIMINARY SEQUENCE FROM N.A.
 RX MEDLINE=91017504; PubMed=1699225;
 RA Adair W.S., Apt K.E.;
 RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs
 encoding cell wall hydroxyproline-rich glycoproteins".;

```

RL  Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
CC  -!- FUNCTION: Major component of the outer cell wall W6 (crystalline)
CC  layer.
CC  -!- SUBUNIT: Associates with GP2 and GP3
CC  -!- PTM: N-glycosylated and O-glycosylated.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AF309494; AAG45420.1; -
DR  EMBL; M58496; AAA69706.1; ALT_SEQ.
DR  GlycoSuiteDB; Q9PQ6; -
DR  InterPro; IPR002965; P-rich_extensin.
DR  PRINTS; PRO1217; PRICHEXTENSIN.
DR  PRINTS; PRO1218; PSTIEXTENSIN.
KW  Glycoprotein; Repeat; Signal.
FT  SIGNAL 1 29 POTENTIAL.
FT  CHAIN 30 555 VEGETATIVE CELL WALL PROTEIN GPI.
FT  DOMAIN 40 339 49 X 5 AA APPROXIMATE PPSXP REPEATS.
FT  DOMAIN 259 279 POLY-PRO.
FT  CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ  SEQUENCE 555 AA; 54219 MW; 6A584A90465502F5 CRC64;

Query Match 15.5%; Score 159.5; DB 1; Length 555;
Best Local Similarity 32.8%; Pred. No. 0.063;
Matches 38; Conservative 6; Mismatches 49; Indels 23; Gaps 3;

QY 33 KPPLPQOPAPANQDNSSQNTLRTPPIP-----APAKPAAPRPDLRSPGVENKLI 88
DB 179 PSPSPVPSPAPP-----SPAPPVPSAPPSPAPPVPPSPAPPSPAPPSP 228

QY 89 PVSQSPASSTPLPDGTGTPNPNRAVTPVSQGSNSSADPKAPPPPPVSSGPP 144
DB 229 PSPAPPSPAPPSPAPPVPPSPAPPSPA-----PPSPKPPAPPSPPPPPPP 275

RESULT 5
ID ENAH_MOUSE STANDARD; PRT; 802 AA.
AC Q03173; P70430; P70431; P70432; P70433;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Enabled protein homolog (NPC derived proline-rich protein 1) (NDPP-1).
GN ENAH OR MENA OR NDPPI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=93041923; PubMed=1420303;
RA Sazuka T., Tomooka Y., Kathju S., Ikawa Y., Noda M., Kumar S.;
RT "Identification of a developmentally regulated gene in the mouse
RL central nervous system which encodes a novel proline rich protein.";
RL Biochim. Biophys. Acta 1132:240-248(1992).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4 AND 5), FUNCTION, AND SUBCELLULAR
RP LOCATION.
RC TISSUE=Brain;
RX MEDLINE=97015079; PubMed=8861907;
RA Gertler F.B., Niebuhr K., Reinhard J., Soriano P.;
RT "Mena, a relative of VASP and Drosophila Enabled, is implicated in the
RT control of microfilament dynamics.";
```

```

RL  Cell 87:227-239(1996).
RN  [3]
RP  FUNCTION, AND SUBUNIT.
RX  MEDLINE=99166867; PubMed=10069337;
RA  Lanier L.M., Gates M.A., Witke W., Menzies A.S., Wehman A.M.,
RA  Macklis J.D., Kwiatkowski D., Soriano P., Gertler F.B.;
RT  "Mena is required for neurulation and commissure formation.";
RL  Neuron 22:313-325(1999).
CC  -!- FUNCTION: May be involved in microfilament assembly and cell
CC  motility. Induces the formation of F-actin rich outgrowths in
CC  fibroblasts. Required for neurulation and commissure formation.
CC  -!- SUBUNIT: Binds profilin.
CC  -!- SUBCELLULAR LOCATION: Localized to focal adhesions and, to a
CC  lesser extent, leading edges and stress fibers.
CC  -!- ALTERNATIVE PRODUCTS:
CC  Event-Alternative splicing; Named isoforms=5;
CC  Name=5; Synonyms=Mena++;
CC  IsoId=Q03173-1; Sequence=Displayed;
CC  Name=1;
CC  IsoId=Q03173-2; Sequence=VSP_007255, VSP_007256;
CC  Name=2; Synonyms=Mena;
CC  IsoId=Q03173-3; Sequence=VSP_007259, VSP_007260;
CC  Name=3; Synonyms=Mena+;
CC  IsoId=Q03173-4; Sequence=VSP_007259;
CC  Name=4; Synonyms=Mena++;
CC  IsoId=Q03173-5; Sequence=VSP_007257, VSP_007258;
CC  -!- TISSUE SPECIFICITY: IN THE HEART AND TESTIS AND LESS SO IN THE
CC  LUNG, SKELETAL MUSCLE, KIDNEY, PANCREAS AND BRAIN.
CC  -!- SIMILARITY: Contains 1 WH1 domain.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; D10727; BAA01570.1; -
DR  EMBL; U72520; AAC52863.1; -
DR  EMBL; U72521; AAC52864.1; -
DR  EMBL; U72522; AAC52865.1; -
DR  EMBL; U72523; AAC52866.1; -
DR  PIR; S27200; S27200.
DR  MGI; 108360; Enah.
DR  InterPro; IPR001960; WH1.
DR  Pfam; PF00568; WH1; 1.
DR  SMART; SM00461; WH1; 1.
KW  Developmental protein; Neurogenesis; Alternative splicing.
FT  DOMAIN 1 108 WH1.
FT  DOMAIN 442 464 POLY-PRO.
FT  DOMAIN 542 552 POLY-PRO.
FT  DOMAIN 562 574 POLY-PRO.
FT  DOMAIN 578 589 POLY-PRO.
FT  DOMAIN 593 605 POLY-PRO.
FT  VARSPLIC 1 412 Missing (in isoform 1).
FT  VARSPLIC 535 631 /FTid=VSP_007255.
FT  VARSPLIC 117 131 Missing (in isoform 4).
FT  VARSPLIC 132 135 /FTid=VSP_007257.
FT  VARSPLIC 117 135 C1FC -> VFVL (in isoform 4).
FT  VARSPLIC 117 135 /FTid=VSP_007258.
FT  VARSPLIC 259 500 Missing (in isoform 2 and isoform 3).
FT  VARSPLIC 802 AA; 85844 MW; 592BB975EE20F77F CRC64;
SQ  SEQUENCE 802 AA; 85844 MW; 592BB975EE20F77F CRC64;
```


Query Match 15.4%; Score 158; DB 1; Length 802;
 Best Local Similarity 26.3%; Pred. No. 0.1;
 Matches 59; Conservative 15; Mismatches 62; Indels 88; Gaps 10;

QY 13 NKTERSTAPLNM-----QISALRNDP-----33
 DB 348 NKNRPSSVNTFSSQPPAAKSCAMPTSNFSLPPSPIMISSPPGKATGPRVLPVCVS 407
 QY 34 KPLPQQPPAPANQDN-----SSQNTRLQPTPIIPAPAKPAAPRPL-----76
 DB 408 SFVQMPSPPTAPNGSLDSVTYVSPFTISGAAPPPPPPPPPPPPPPPPLPPLPLAS 467
 QY 77 -----DRSPGVENKLI-----PSVGSPPASS-TPLPPD-GTGNSTPNNRAY- 116
 DB 468 LSHSGSQASPPPGTTLASTPSSKPSVLPSPSACAPASAEETPLNPELGDSSASEPGLQAS 527
 QY 117 ----TFVQSGSNSSADPKAPPPPPVSSGE-----PPTLGNP 150
 DB 528 QPAESPTFG--LVLGPPAPPPPPLPLSPGAYASALPPPPPGPPP 569

RESULT 6
 APG_BRANA STANDARD; PRT; 449 AA.
 AC P40603;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Anter-specific proline-rich protein APG (Protein CEX) (Fragment).
 GN APG OR CEX.
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94004980; PubMed=8401599;
 RA Roberts M.R., Foster G.D., Blundell R.P., Robinson S.W., Kumar A.,
 RA Draper J.J., Scott R.J.
 RT "Gametophytic and sporophytic expression of an anther-specific
 Arabidopsis thaliana gene.";
 RL Plant J. 3:111-120(1993).
 CC -!- TISSUE SPECIFICITY: FOUND IN ANTHER, ONLY IN MALE FERTILE PLANTS.
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN MALE GAMETOGENESIS. DURING
 CC MICROSPORE DEVELOPMENT, HIGHER EXPRESSION IS FOUND DURING
 CC MICROSPORE MITOSIS WITH A DRAMATIC DECLINE DURING POLLEN
 CC MATURATION.
 CC -!- SIMILARITY: BELONGS TO THE "GDSL" FAMILY OF LIPOLYTIC ENZYMES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X60376; CAA42924.1; -;
 DR PIR; S16748; S16748.
 DR InterPro; IPR001087; Lipase_GDSL.
 DR InterPro; IPR003882; Pistil_extensin.
 DR Pfam; PF00657; Lipase_GDSL; 1.
 DR PRINTS; PR01218; PSTLEXENSIN.
 DR PROSITE; PS01098; LIPASE_GDSL_SER; 1.
 FT NON_TER 1
 FT ACT_SITE 132 132 BY SIMILARITY.
 FT ACT_SITE 428 428 POTENTIAL.
 FT ACT_SITE 428 428
 SQ SEQUENCE 449 AA; 48779 MW; 9EFB6A3AB28EA15 CRC64;

Query Match 15.2%; Score 156; DB 1; Length 449;
 Best Local Similarity 33.3%; Pred. No. 0.079;

Matches 39; Conservative 9; Mismatches 57; Indels 12; Gaps 3;
 QY 33 KPPLPQQPPAPANQDNSSQNTRLQPTPIIPAPAKPAAPRPLDRSPGVENKLI PSVG 92
 DB 2 PXPQKPPPKPQKPPAPTPSPCPKPPKPPAPTPSPCPKPPKPPAPTPSPCPKPPKPPAPG 61
 QY 93 -----SPASSTPLPDGTGTGNSTPNNRAYTPVSGSNSSADPKAPPPPPVSSGPP 144
 DB 62 PSPKPGSPSPKPPSPAPKVP-----PSPSPKPPKPPAPSPAP-SPKPP 111

RESULT 7
 WAIP_HUMAN STANDARD; PRT; 503 AA.
 AC O43516; Q15220; Q9UNP1;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Wiskott-Aldrich syndrome protein interacting protein (WASP interacting
 DE protein) (PRPL-2 protein).
 GN WASPIP OR WIP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98070810; PubMed=9405671;
 RA Ramesh N., Anton I.M., Hartwig J.H., Geha R.S.;
 RT "WIP, a protein associated with Wiskott-Aldrich syndrome protein,
 RT induces actin polymerization and redistribution in lymphoid cells.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:14671-14676(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Tonsil;
 RA Kreideweiss S., Delany-Heiken P., Nordheim A., Ruhlmann A.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 192-503 FROM N.A., AND VARIANT ALA-495.
 RX MEDLINE=99218549; PubMed=10202051;
 RA Stewart D.M., Tian L., Nelson D.L.;
 RT "Mutations that cause the Wiskott-Aldrich syndrome impair the
 RT interaction of Wiskott-Aldrich syndrome protein (WASP) with WASP
 RT interacting protein.";
 RL J. Immunol. 162:5019-5024(1999).
 CC -!- FUNCTION: MAY HAVE DIRECT ACTIVITY ON THE ACTIN CYTOSKELETON.
 CC INDUCES ACTIN POLYMERIZATION AND REDISTRIBUTION.
 CC -!- SUBUNIT: BINDS TO WASP, PROFILIN AND ACTIN.
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PERIPHERAL BLOOD
 CC MONONUCLEAR CELLS, SPLEEN, PLACENTA, SMALL INTESTIN, COLON,
 CC THYMUS. LOWER EXPRESSION IN OVARY, HEART, BRAIN, LUNG, LIVER,
 CC SKELETAL MUSCLE, KIDNEY, PANCREAS, PROSTATE, TESTIS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF031588; AAC03767.1; -;
 DR EMBL; AF106062; AAD45972.1; -;
 DR EMBL; X86019; CAA60014.1; -;
 DR Genew; HGNC:12736; WASPIP.
 DR MIM; 602357; -;
 DR GO; GO:0015629; C:actin cytoskeleton; TAS.
 DR GO; GO:0003779; F:actin binding activity; TAS.
 DR GO; GO:0005522; F:profilin binding activity; TAS.
 DR GO; GO:0008154; P:actin polymerization and/or depolymerization; TAS.
 DR GO; GO:0006461; P:protein complex assembly; TAS.
 DR InterPro; IPR003124; WH2.
 DR Pfam; PF02205; WH2; 1.


```

RA Numasawa T., Koga H., Ueyama K., Maeda S., Sakou T., Harata S.,
RA Leppert M., Inoue I.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Corella A., Vergara A., Paez G., de Miguel C., Encio I.;
RT "Molecular cloning and characterization of the human HRXR gene and 5'
RT flanking region.";
RN Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RL [6]
RP SEQUENCE FROM N.A.
RA Tubby B.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: INVOLVED IN RETINOIC ACID RESPONSE PATHWAY. BINDS 9-CIS
CC RETINOIC ACID (9C-RA).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=Long;
CC IsoId=P28702-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P28702-2; Sequence=Not described;
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: Belongs to the nuclear hormone receptor family. NR2
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X63522; CAA45087.1; -
CC EMBL; M84820; AAA60293.1; -
CC EMBL; AF065396; AAC18599.1; -
CC EMBL; AF120161; AAD13794.1; -
CC EMBL; AL031228; CAA20239.1; -
CC EMBL; BC001167; AAH01167.1; -
CC PIR; S37781; S37781.
CC PDB; 1H90; 22-MAY-02.
CC TRANSFAC; T01334; -
CC DR Genew; HGNC:10478; RXRD.
CC DR MIM; 180246; -
CC DR GO; GO:0004886; F:retinoid-X receptor activity; TAS.
CC DR GO; GO:0003713; F:transcription co-activator activity; TAS.

```

```

DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Studhm_receptor.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR PRODOM; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Multigene family; Alternative splicing; 3D-structure.
FT DOMAIN 1 204 MODULATING (BY SIMILARITY).
FT DNA_BIND 205 270 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 205 225 C4-TYPE.
FT ZN_FING 241 265 C4-TYPE.
FT DOMAIN 271 330 HINGE.
FT DOMAIN 331 533 LIGAND-BINDING (BY SIMILARITY).
FT DOMAIN 61 198 PRO-RICH.
FT CONFLICT 112 112 S -> T (IN REF. 2).
SQ SEQUENCE 533 AA; 56921 MW; D0069FE93AC16A04 CRC64;

Query Match 14.6%; Score 150.5; DB 1; Length 533;
Best Local Similarity 31.0%; Pred. No. 0.17;
Matches 40; Conservative 11; Mismatches 33; Gaps 5;

QY 31 NDPRLPQ--OPPAPANQDNSSONTRLQTPPTPAPA-----PKPAAPPRLDRESP 81
DB 86 SSPNPLPQGVPPPPPG-----PPLPPTAFSLGSGGAPPPPPPPPLGSPFP 134
QY 82 GVENKLIPSGVSPASSTPLPDGTPNNTNRAVTPVSGSNSSSSADPKAP----- 133
DB 135 VISS-----SMGSPGLPPAPPFGFSGVSPQINSTVSLPGGGSGPPEDYKPPVLGVRLH 190
QY 134 -PPPPVSSG 141
DB 191 CPPPPGGPG 199

RESULT 10
MUC2_HUMAN
ID MUC2_HUMAN STANDARD; PRT; 5179 AA.
AC Q02817; Q14878;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mucin 2 precursor (Intestinal mucin 2).
GN MUC2 OR SMUC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R., Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor.";
RL J. Biol. Chem. 269:2440-2446(1994).
[2]
RN SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R., Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region.";
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.

```

RX MEDLINE-91358717; PubMed=1885763;
 RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
 RA Petersen G.M., Kim Y.S.;
 RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
 RL J. Clin. Invest. 88:1005-1013(1991).
 CC -!- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
 CC OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
 CC PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS
 CC AGENTS AT MUCOSAL SURFACES.
 CC -!- SUBUNIT: MULTIMERIC.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
 CC BRONCHUS, CERVIX AND GALL BLADDER.
 CC -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
 CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
 CC -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
 CC VARIES AMONG DIFFERENT ALLELES.
 CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
 CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
 CC -!- SIMILARITY: Contains 2 VWFC domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L21998; AB95295.1; -;
 CC EMBL; M74027; AA59675.1; -;
 CC EMBL; M94131; AA59163.1; -;
 CC EMBL; M94132; AA59164.1; -;
 CC PIR; A49963; A43932.
 CC Genew: HGNC:7512; MUC2.
 CC MIN; 158370; -;
 CC GO; GO:0005803; C:secretory vesicle; TAS.
 CC InterPro; IPR006208; Cys_knot.
 CC InterPro; IPR006207; Cys_knot_C.
 CC InterPro; IPR006209; EGF-like.
 CC InterPro; IPR002919; TIL_Cysrich.
 CC InterPro; IPR001007; VWFC_C.
 CC InterPro; IPR001846; VWFC_D.
 CC Pfam; PF00007; Cys_knot; 1.
 CC Pfam; PF01826; TIL; 1.
 CC Pfam; PF00093; vwc; 1.
 CC Pfam; PF00094; vwd; 4.
 CC SMART; SM00214; VWC; 2.
 CC SMART; SM00216; VWD; 4.
 CC PROSITE; PS01185; CTCK_1; 1.
 CC PROSITE; PS01225; CTCK_2; 1.
 CC PROSITE; PS00022; EGF_1; UNKNOWN_1.
 CC PROSITE; PS01208; VWFC_1; 2.
 CC PROSITE; PS0184; VWFC_2; 2.
 KW Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 5179 MUCIN 2.
 FT DOMAIN 1401 1747 APPROXIMATE REPEATS.
 FT REPEAT 1401 1416 1.
 FT REPEAT 1417 1432 2.
 FT REPEAT 1433 1448 3.
 FT REPEAT 1449 1464 4.
 FT REPEAT 1465 1471 5.
 FT REPEAT 1472 1478 6.
 FT REPEAT 1479 1494 7A.
 FT REPEAT 1495 1517 7B.
 FT REPEAT 1518 1533 8A.
 FT REPEAT 1534 1556 8B.
 FT REPEAT 1557 1572 9A.
 FT REPEAT 1573 1596 9B.
 FT REPEAT 1597 1612 10A.
 FT REPEAT 1613 1635 10B.

FT REPEAT 1636 1651 11A.
 FT REPEAT 1652 1675 11B.
 FT REPEAT 1676 1683 12.
 FT REPEAT 1684 1699 13.
 FT REPEAT 1700 1715 14.
 FT REPEAT 1716 1731 15.
 FT REPEAT 1732 1747 16.
 FT DOMAIN 4815 4886 VWFC 1.
 FT DOMAIN 4924 4991 VWFC 2.
 FT DOMAIN 5075 5160 CTCK.
 FT DISULFID 5075 5122 BY SIMILARITY.
 FT DISULFID 5089 5136 BY SIMILARITY.
 FT DISULFID 5098 5152 BY SIMILARITY.
 FT DISULFID 5102 5154 BY SIMILARITY.
 FT DISULFID ? 5159 BY SIMILARITY.
 FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 770 770 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 894 894 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1154 1154 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1215 1215 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1230 1230 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1246 1246 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1787 1787 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 1820 1820 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 4339 4339 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 4351 4351 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 4362 4362 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 4373 4373 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 4422 4422 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 4438 4438 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 4502 4502 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 4616 4616 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 4627 4627 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 4752 4752 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 4787 4787 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 4881 4881 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 4888 4888 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 4955 4955 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 4970 4970 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 5019 5019 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 5038 5038 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 5069 5069 N-LINKED (GLCNAC. . .)
 FT CONFLICT 1351 1351 H -> L (IN REF. 3).
 FT CONFLICT 1412 1412 T -> S (IN REF. 3).
 FT CONFLICT 1449 1449 L -> P (IN REF. 3).
 FT CONFLICT 1504 1504 M -> T (IN REF. 3).
 FT CONFLICT 4192 4192 G -> S (IN REF. 2).
 SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;

Query Match 14.6%; Score 150; DB 1; Length 5179;
 Best Local Similarity 25.6%; Pred. No. 1.3;

Matches 41; Conservative 20; Mismatches 65; Indels 34; Gaps 5;

QY 15 TERSTAPLNTQISALRNDPPLFCQPPAPANQONSSQNTLRQPTTPIAPAPKPAAP-- 72
 Db 1428 TTTTPPTTTPSPPTTTTTLPLTTPSPPTTTPPTTTPPTTTPPTTTPSP 1486
 QY 73 -----PRPLDRESGKVENKLIPIVSGSPASSPLPPDGT-----GNSTPNR 114
 Db 1487 PTTTTPPTTTPSPPTTTPPTTTP-----PASTTTLPTTTPSPPTTTPPTTTPSP 1542
 QY 115 AVTPVSGQSSSSADPKAPP-----PPPVSSGEPPT 145
 Db 1543 TTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPSPPT 1582

RESULT 11
 EXLP_TOBAC

ID EXLP_TOBAC STANDARD; PRT; 426 AA.
 AC Q03211;

RX MEDLINE-20020275; PubMed-10551867;
 RA Zitzer H., Hoenck H.-H., Baechner D., Richter D., Kreienkamp H.-J.;
 RT "Somatostatin receptor interacting protein defines a novel family of
 RT multidomain proteins present in human and rodent brain.";
 RL J. Biol. Chem. 274:32997-33001(1999).
 RN [6]
 RP INTERACTION WITH HOMER-1, AND SUBCELLULAR LOCATION.
 RX MEDLINE-99360651; PubMed-10433269;
 RA Tu J.-C., Xiao B., Naisbitt S., Yuan J.-P., Petralia R.S., Brakeman P.,
 RA Doan A., Aakalu V.K., Lanahan A.A., Sheng M., Worley P.F.;
 RT "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of
 RT postsynaptic density proteins.";
 RL Neuron 23:583-592(1999).
 RN [7]
 RP INTERACTION WITH SPTAN1.
 RX MEDLINE-21523912; PubMed-11509555;
 RA Bockers T.M., Maneza M.G., Kretz M.R., Bockmann J., Weise C.,
 RA Buck F., Richter D., Gundelfinger E.D., Kreienkamp H.-J.;
 RT "Synaptic scaffolding proteins in rat brain. Ankyrin repeats of the
 RT multidomain Shank protein family interact with the cytoskeletal
 RT protein alpha-fodrin.";
 RL J. Biol. Chem. 276:40104-40112(2001).
 RN [8]
 RP FUNCTION.
 RX MEDLINE-21389514; PubMed-11498055;
 RA Sala C., Piech V., Wilson N.R., Passafaro M., Liu G., Sheng M.;
 RT "Regulation of dendritic spine morphology and synaptic function by
 RT Shank and Homer.";
 RL Neuron 31:115-130(2001).
 RN [9]
 RP REVIEW.
 RX MEDLINE-20267967; PubMed-10806096;
 RA Sheng M., Kim E.;
 RT "The Shank family of scaffold proteins.";
 RL J. Cell Sci. 113:1851-1856(2000).
 CC -!- FUNCTION: Seems to be an adapter protein in the postsynaptic
 CC density (PSD) of excitatory synapses that interconnects receptors
 CC of the postsynaptic membrane including NMDA-type and metabotropic
 CC glutamate receptors, and the actin-based cytoskeleton. May play a
 CC role in the structural and functional organization of the
 CC dendritic spine and synaptic junction. Overexpression promotes
 CC maturation of dendritic spines and the enlargement of spine heads
 CC via its ability to recruit Homer to postsynaptic sites, and
 CC enhances presynaptic function.
 CC -!- SUBUNIT: May homomultimerize via its SAM domain. Interacts with
 CC SPTAN1, Homer-1 and DLGAP1/GKAP. Is part of a complex with
 CC DLG4/PSD-95 and DLGAP1/GKAP. Interacts with SSTR2 C-terminus via
 CC the PDZ domain (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of
 CC neuronal cells. Colocalizes with alpha-latrotoxin receptor 1.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms-5;
 CC Comment-Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=Q9WV48-1; Sequence-Displayed;
 CC Name=2;
 CC IsoId=Q9WV48-2; Sequence-VSP_006072, VSP_006073;
 CC Name=3;
 CC IsoId=Q9WV48-3; Sequence-VSP_006074;
 CC Name=4; Synonyms=A;
 CC IsoId=Q9WV48-4; Sequence-VSP_006075;
 CC Name=5;
 CC IsoId=Q9WV48-5; Sequence-VSP_006076, VSP_006077;
 CC -!- TISSUE SPECIFICITY: Expressed only in brain (neuropil of cortex,
 CC CA1 region hippocampus and molecular layer of cerebellum).
 CC -!- DEVELOPMENTAL STAGE: Expression increases from low levels at birth
 CC to high levels at 3-4 weeks before dropping slightly in adulthood.
 CC Expressed in the cortex and the molecular layer of the cerebellum
 CC at postnatal day 7. Isoform 2 expression does not change during
 CC development of both cortex and cerebellum. Isoform 4 expression
 CC decreases significantly during development of cortex but not
 CC cerebellum.
 CC -!- SIMILARITY: BELONGS TO THE SHANK FAMILY.

CC -!- SIMILARITY: Contains 7 ANK repeats.
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
 CC -!- SIMILARITY: Contains 1 SH3 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF102855; RAD04569.2;
 DR EMBL; AF131951; RAD29417.1; ALT_INIT.
 DR EMBL; AF159046; RAD42975.1;
 DR EMBL; AF141904; AAF02498.1; ALT_INIT.
 DR HSSP; P00519; IABL.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001660; SAM.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00023; ank; 7.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00536; SAM; 1.
 DR Pfam; PF00018; SH3; 1.
 DR ProDom; PD000066; SH3; 1.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00228; PDZ; 1.
 DR SMART; SM00454; SAM; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 3.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50106; PDZ; 1.
 DR PROSITE; PS50002; SH3; 1.
 DR PROSITE; PS50105; SAM_DOMAIN; 1.
 KW ANK repeat; SH3 domain; Repeat; Alternative splicing.
 FT REPEAT 195 210 ANK 1.
 FT REPEAT 212 245 ANK 2.
 FT REPEAT 246 278 ANK 3.
 FT REPEAT 279 312 ANK 4.
 FT REPEAT 313 345 ANK 5.
 FT REPEAT 346 378 ANK 6.
 FT REPEAT 379 395 ANK 7.
 FT DOMAIN 554 613 SH3.
 FT DOMAIN 663 757 PDZ.
 FT DOMAIN 2104 2167 SAM.
 FT DOMAIN 929 932 POLY-PRO.
 FT DOMAIN 1010 1015 POLY-HIS.
 FT DOMAIN 1022 1027 POLY-HIS.
 FT DOMAIN 1194 1199 POLY-GLY.
 FT DOMAIN 1850 1860 POLY-PRO.
 FT VARSPPLIC 1 613 Missing (in isoform 2).
 FT VARSPPLIC 615 654 /FtId-VSP_006072.
 FT VARSPPLIC 654 SOEGROESRDKAKELFRHYTVGSYDSDAPSLDGDSDG
 FT VARSPPLIC 646 -> MALSAVGSGPGGALPQPPALSSSWPALGPRRRSVWY
 FT VARSPPLIC 797 804 /FtId-VSP_006073.
 FT VARSPPLIC 1930 1943 /FtId-VSP_006074.
 FT VARSPPLIC 1944 2167 /FtId-VSP_006075.
 FT VARSPPLIC 1944 2167 /FtId-VSP_006076.
 FT VARSPPLIC 1944 2167 /FtId-VSP_006077.
 FT CONFLICT 1141 1141 S -> T (IN REF. 1).
 FT CONFLICT 1174 1174 S -> N (IN REF. 2).
 FT CONFLICT 1246 1246 R -> K (IN REF. 1).
 FT CONFLICT 1323 1323 A -> T (IN REF. 1).
 FT CONFLICT 1331 1331 S -> D (IN REF. 1).
 FT CONFLICT 1726 1726 S -> N (IN REF. 2).

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:33 ; Search time 112.475 Seconds
(without alignments)
445.097 Million cell updates/sec

Title: US-09-915-543-15_COPY_199_392
Perfect score: 1028
Sequence: 1 TIVSFHQNISNNKTERSTA.....LFPDEKFTGAQSGGQQNP 194

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.rodent:*
- 12: sp.virus:*
- 13: sp.vertebrate:*
- 14: sp.unclassified:*
- 15: sp.virus:*
- 16: sp.bacteriaph:*
- 17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	177.5	17.3	473	10	Q39620	Q39620 chlamydomon
2	168.5	16.4	876	3	Q9P4Y9	Q9P4Y9 neurospora
3	167.5	16.3	1269	10	Q8W5K6	Q8W5K6 oryza sativ
4	163	15.9	1307	10	Q9LVN1	Q9LVN1 arabidopsis
5	162.5	15.8	574	3	Q36027	Q36027 schizosacch
6	162.5	15.8	1011	3	Q9P944	Q9P944 pneumocysti
7	162.5	15.8	1151	13	Q57580	Q57580 gallus gall
8	161.5	15.7	816	4	Q96G51	Q96G51 homo sapien
9	160.5	15.6	2321	12	Q9DGT6	Q9DGT6 turkey herp
10	160	15.6	488	16	Q8DGUI	Q8DGUI synecococc
11	160	15.6	839	16	Q9RX57	Q9RX57 deinococcus
12	158.5	15.4	351	10	Q39492	Q39492 chlamydomon
13	158.5	15.4	420	5	Q9VZC2	Q9VZC2 drosophila
14	158.5	15.4	1188	10	Q41805	Q41805 zea mays (m
15	158	15.4	802	11	P70433	P70433 mus musculu
16	157.5	15.3	464	10	Q41645	Q41645 volvox cart

17	156	15.2	1315	10	Q9SPM0	Q9SPM0 zea mays (m
18	155	15.1	763	2	Q9XDH2	Q9XDH2 mycobacteri
19	154.5	15.0	386	10	Q9FPQ5	Q9FPQ5 chlamydomon
20	154.5	15.0	731	10	Q65530	Q65530 arabidopsis
21	153.5	14.9	326	10	O22514	O22514 santalum al
22	153.5	14.9	637	4	Q9BSV4	Q9BSV4 homo sapien
23	153	14.9	493	11	Q8XLI7	Q8XLI7 mus musculu
24	152.5	14.8	379	10	Q9SUX2	Q9SUX2 arabidopsis
25	152	14.8	883	2	Q9RF11	Q9RF11 myxococcus
26	152	14.8	956	10	Q9LJ64	Q9LJ64 arabidopsis
27	151.5	14.7	602	2	Q9AKP3	Q9AKP3 rickettsia
28	151.5	14.7	964	5	Q26963	Q26963 trypanosoma
29	151	14.7	616	4	Q9HGK5	Q9HGK5 homo sapien
30	151	14.7	1997	10	Q8LRM7	Q8LRM7 chlamydomon
31	150.5	14.6	598	16	Q8VKN7	Q8VKN7 mycobacteri
32	150.5	14.6	699	11	Q8VIJ6	Q8VIJ6 mus musculu
33	150	14.6	309	12	Q8U2B4	Q8U2B4 grapevine f
34	150	14.6	409	10	Q9S8M1	Q9S8M1 volvox cart
35	150	14.6	556	4	O15411	O15411 homo sapien
36	150	14.6	1157	4	Q96JK7	Q96JK7 homo sapien
37	150	14.6	3124	4	Q96L91	Q96L91 homo sapien
38	149	14.5	188	11	Q62106	Q62106 mus musculu
39	149	14.5	302	5	Q9GZH1	Q9GZH1 caenorhabdi
40	149	14.5	712	10	Q8RWX5	Q8RWX5 arabidopsis
41	149	14.5	1201	10	Q9C6S1	Q9C6S1 arabidopsis
42	148.5	14.4	273	10	Q9ARY7	Q9ARY7 oryza sativ
43	148.5	14.4	731	10	Q9ZUE0	Q9ZUE0 arabidopsis
44	148.5	14.4	752	10	Q9LQC5	Q9LQC5 arabidopsis
45	148.5	14.4	1202	4	Q8WXE0	Q8WXE0 homo sapien

ALIGNMENTS

RESULT 1

Q39620 PRELIMINARY; PRT: 473 AA.
AC Q39620;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 21, Last annotation update)
DE VSP-3 protein precursor.
GN VSP-3.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CC-621;
RX MEDLINE=95093034; PubMed=8000007;
RA Woessner J.P., Molendijk A.J., van Egmond P., Klis F.M.,
RA Goodenough U.W., Haring M.A.;
RT "Domain conservation in several volvoclean cell wall proteins.";
RL Plant Mol. Biol. 26:947-960(1994).
DR EMBL; L29029; AA053953.1; -;
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
KW Signal.
FT SIGNAL. 1 31 POTENTIAL.
FT CHAIN 32 473 POTENTIAL.
SQ SEQUENCE 473 AA; 47532 MW; 38506131FAA674A4 CRC64;

Query Match 17.3%; Score 177.5; DB 10; Length 473;
Best Local Similarity 30.5%; Pred. No. 1.7e-05;
Matches 46; Conservative 19; Mismatches 59; Indels 27; Gaps 5;

QY 14 KTERETALNQISALRNDPKLPQQPAPANQDNSSQNRLOPTPIIPAPKPAAPP 73
Db 330 KASPSPSPSVQPAKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 376
QY 74 RPLDRESFCVENKLIPSVGSPASSPPLPDGTGNSTPN-NRAVTPVSGQSNSSADPKA 132

RESULT 7		
OS7580		
ID	OS7580	PRELIMINARY; PRT; 1151 AA.
AC	OS7580;	
DT	01-JUN-1998 (TREMBlrel. 06, Created)	
DT	01-JUN-1998 (TREMBlrel. 06, Last sequence update)	
DT	01-MAR-2003 (TREMBlrel. 23, Last annotation update)	
DE	High molecular mass nuclear antigen (fragment).	
OS	Gallus gallus (Chicken).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;	
CC	Gallus.	
OX	NCBI_TaxID=9031;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=98034404; PubMed=9365273;	
RA	Shimada K., Harata M., Mizuno S.;	
RT	"A nuclear matrix-associated high molecular mass nuclear antigen,	

RT HMNA, of chicken and marked decrease of its immunoreactivity during
 RT the progression of s phase."
 RL J. Cell Sci. 110:3031-3041(1997).
 DR EMBL: D8440; BAA24137.1; -.
 DR InterPro: IPR002965; P_rich_extensn.
 DR PRINTS: PR01217; PRICHEXTENS.
 FT NON_TER 1151
 FT SEQUENCE 1151 AA; 109708 MW; 2403F6835F9A2AB3 CRC64;
 SQ
 Query Match 15.8%; Score 162.5; DB 13; Length 1151;
 Best Local Similarity 33.1%; Pred. No. 0.00052;
 Matches 45; Conservative 12; Mismatches 40; Indels 39; Gaps 6;
 QY 33 PKPLPQQPAPANQDSSQNTRLQPTP-PAPAPAPKPAAPPRP-----LDRESGVE 84
 DB 3 PTPPPRNPTP-----PPAPAPAPAPAPAPPRKWPVIAELHFAAPQPP 49
 QY 85 NKLPISVSGSPASSTPLPDGTGPNSTPNNRAVTPVSGSSSSADPKAP---PPPPVSSG 141
 DB 50 PKWPIGGAP-----PPGTGP-----TPPSKPTGDGAAPKASBELTSPPPASP 94
 QY 142 EPTLTGENPDGLSQBQ 157
 DB 95 SPDPGKAPSGAGEAE 110
 RESULT 8
 Q96G51 PRELIMINARY; PRT; 816 AA.
 ID Q96G51; Q96G91;
 AC Q96G51-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-DEC-2003 (Tremblrel. 23, Last annotation update)
 DE Mitogen-activated protein kinase 7.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas, and Muscle;
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RA Strausberg R.;
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 CC -I- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL: BC009963; AAH09963.1; -.
 DR EMBL: BC007992; AAH07992.1; -.
 DR EMBL: BC007404; AAH07404.1; -.
 DR EMBL: BC030134; AAH30134.1; -.
 DR HSSP: P24941; IBUH.
 DR InterPro: IPR003527; MAP_kin.
 DR InterPro: IPR000719; Prot_kinase.
 DR InterPro: IPR002965; P_rich_extensn.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR PRINTS: PR01217; PRICHEXTENS.
 DR ProDom: PD000001; Prot_kinase; 1.
 DR PROSITE: PS01351; MAPK; 1.
 DR PROSITE: PS0107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
 SQ SEQUENCE 816 AA; 88386 MW; 27729FE31658CE3B CRC64;
 Query Match 15.7%; Score 161.5; DB 4; Length 816;
 Best Local Similarity 32.1%; Pred. No. 0.00044;
 Matches 54; Conservative 7; Mismatches 64; Indels 43; Gaps 7;

QY 17 RSTAPLNTQISALRNDPKLP-----QQPAPANQDSSQNTRLQPTTPIP 63
 DB 577 RPAAPALTSVPAPAPAPTPTPTVPTSPPPGFVAQPTGPPQASGTSVPPQACPPP 636
 QY 64 APAPKPAAPRPLDRESQ-----VENKLIPSVGSPASSTP-----LPPDGT 105
 DB 637 GPAPHTGPPGPIPVAPPAQIATSTSLAAQSLVPPGPGSGSTPGVLPPPPGLPPDA 696
 QY 106 G--PNS-----TPNNRAVTPVSGSSSSADPKAPPPPPVSSGEPPTLG 147
 DB 697 GGAQSSMESPDVNLVT--QQLSKSQVEDPL-----PPVFGTPKGSG 738
 RESULT 9
 Q9DGT6 PRELIMINARY; PRT; 2321 AA.
 ID Q9DGT6
 AC Q9DGT6
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE RSL immediate-early gene transactivator ICP4-like protein (RSL
 DE immediate-early gene transactivator-like protein).
 GN MDV084 OR MDV100.
 OS Turkey herpesvirus.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Marek's disease-like viruses.
 OX NCBI_TaxID=10390;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD5;
 RX MEDLINE=20392152; PubMed-10933706;
 RA Tulman E.R., Afonso C.L., Lu Z., Zsak L., Rock D.L., Kutish G.F.;
 RT "The genome of a very virulent Marek's disease virus."
 RL J. Virol. 74:7980-7988(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MD5;
 RA Tulman E.R., Afonso C.L., Lu Z., Zsak L., Rock D.L., Kutish G.F.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF243438; AAG14284.1; -.
 DR EMBL: AF243438; AAG14273.1; -.
 DR InterPro: IPR005205; Herpes_ICP4_C.
 DR InterPro: IPR005206; Herpes_ICP4_N.
 DR InterPro: IPR002965; P_rich_extensn.
 DR Pfam: PF03585; Herpes_ICP4_C; 1.
 DR Pfam: PF03584; Herpes_ICP4_N; 1.
 DR PRINTS: PR01217; PRICHEXTENS.
 DR SEQUENCE 2321 AA; 252959 MW; EF130EB0C4503003 CRC64;
 Query Match 15.6%; Score 160.5; DB 12; Length 2321;
 Best Local Similarity 32.4%; Pred. No. 0.0015;
 Matches 60; Conservative 21; Mismatches 53; Indels 51; Gaps 13;
 QY 12 NKKTERSTAPLNTQISALRNDPKLP-----PQPPP-APANQD-----QNSSQNTL- 56
 DB 241 NRASERQLA--DTAASALR-APSPVFWSPAFDSRYPHLAPANQSDPLCPETSTASAQL 297
 QY 57 ---QPTPTPI---PAPAPKPAAP---PRPLDRESQGVENKLIPSVGSPAS-----ST 98
 DB 298 HTNSFTPTTSTPAPISPTQPPACLPSPAPISVQPPQALLPLIFPTVPEFIQPSF 357
 QY 99 PLPPDGTGP-----NSTPNNRAVTPVSGSSSS-----SADPKAP-PPPVSSGEPPTL 146
 DB 358 PSPQAPSPPAHSSSSCSPLAPSLSSFLSPQLSPAPVSPSPSPPLSPGE---L 414
 QY 147 GENPD 151
 DB 415 APSFD 419
 RESULT 10
 Q8DGU1 PRELIMINARY; PRT; 488 AA.
 ID Q8DGU1

Matches	52;	Conservative	13;	Mismatches	79;	Indels	22;	Gaps	7;
QY	33	KPK-LPQPAPAPANQDNSSONTRLQP---TPPIDAPA---PKPAAP-PRPLDRESPGVE	84						
Db	222	PTPNAPAQTPATPATQAAPAAQTPTAQAATQTPATPAPAAQAQFAGCAPSPAPAAQANAPA	281						
QY	85	NKLIPVSGSPASSPLPDGDTGNSTPNNRAVTVPVSGS-NSSADPKAPPPPPVSSGEP	143						
Db	282	GSVVPEATVPESSTPAAPSAGTPTPTRETAAQTASPAAPNSSAAAPNEASEPV-ACRP	340						
QY	144	PTLGENDDGLSQEOLEHRSLOTLRDIQRMLFPDEKEFTGAQGG	189						
Db	341	GTAASSPESASPVTTPRGET-----PDTRASACTPSAG	374						
 RESULT 12									
Q39492									
ID	Q39492	PRELIMINARY;	PRT;	351	AA.				
AC	Q39492;								
DT	01-NOV-1996	(TREMBLrel. 01, Created)							
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)							
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)							
DE	WP6	protein precursor.							
GN	WP6.								
OS	Chlamydomonas eugametos.								
OC	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;								
OC	Chlamydomonadaceae; Chlamydomonas.								
OX	NCBI_TaxID=3053;								
RN	[]								
RP	SEQUENCE FROM N.A.								
RC	STRAIN=UTEX 10;								
RC	MEDLINE=95093034; PubMed=8000007;								
RA	Goessner J.P., Molendijk A.J., van Egmond P., Klis F.M.,								
RA	Goodenough U.W., Haring M.A.;								
RT	"Domain conservation in several volvoclean cell wall proteins.";								
RL	Plant Mol. Biol. 26:947-960(1994).								
DR	EMBL; L23028; AAB53954.1; --								
DR	InterPro; IPR002965; Prich.extensn.								
DR	PRINTS; PR01217; PRICHEXTENS.								
KW	Signal.								
FT	SIGNAL	1	31	POTENTIAL.					
FT	CHAIN	32	351	POTENTIAL.					
SQ	SEQUENCE	351	AA;	35310	MW;	70AD2EFFC74BB68	CRC64;		
Query Match	15.4%;	Score	158	5;	DB	10;	Length	351;	
Best Local Similarity	30.9%;	Pred	No.	0.00032;					
Matches	46;	Conservative	15;	Mismatches	71;	Indels	17;	Gaps	5;
QY	9	NISNNKTERETAINTQ-----ISALRNDRPKLPQQPPAPANQDNS-----SQNTRLQPT	59						
Db	144	SVYNICIDITRAPYNCSTFNVTIVITTPSPSPSPSPSPSPSPSPSPSPSPSKASPS	203						
QY	60	P-PIPAPAKPAAPPRLDRESGVENKLLIPSVGSPASSPLPDGDTGNSTPNNRVTP	118						
Db	204	PSPKASPSPKASPAAPSPQSPSTPSPKASPAVASQOSPSPSRPSPPTPTP-----SP	258						
QY	119	VSGSNSSAADPKAPP--PPPVSSGEPTT	145						
Db	259	SPRASPPSPASPSASLSLKVSPTSPTT	287						
 RESULT 13									
Q39ZC2									
ID	Q39ZC2	PRELIMINARY;	PRT;	420	AA.				
AC	Q39ZC2;								
DT	01-MAY-2000	(TREMBLrel. 13, Created)							
DT	01-OCT-2002	(TREMBLrel. 22, Last sequence update)							
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)							
DE	CG15021	protein (REI17165p).							
GN	CG15021.								
OS	Drosophila melanogaster (Fruit fly).								
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;								
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha:								

[illegible]

```
Query Match 15.4%; Score 158.5; DB 10; Length 1188;
Best Local Similarity 32.0%; Pred. NO. 0.0011;
Matches 49; Conservative 17; Mismatches 68; Indels 19; Gaps 7;

QY 14 KTERSTAPLNTQISALRNDPKPLP--QPPAPANODQNSSQNTRLQP---TPPIAPAPAK 68
Db 961 KSSPPAPVNLPPPEVKSPPTPVSSPPAPKSSPPAPMSPPPPEVKSPPPAPVSS 1020
QY 69 P-----AAPPRLDRESGVENKLIPIVSGSPASSTPLP---PDGTGPNSTNNRAVTPV 119
Db 1021 PPPPVKSPPPAPVSPPPPVKS---PPPPAPVSPPPPVKSPPPPPAPISSPPPVKSP 1077
QY 120 SGGSSNSSADP-KAPPPP-PVSGEPTLGENP 150
Db 1078 PPAVSSPPPPVKSPPPPAPVSSPPPIKSPPP 1110

RESULT 15
P70433 PRELIMINARY; PRT; 802 AA.
ID P70433 AC P70430; P70431; P70432;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE ENABLED homolog.
GN ENAH OR MENA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=BRAIN;
RA MEDLINE=97015079; PubMed=8861907;
RT Gertler F.B., Niebuhr K., Reinhard M., Wehland J., Soriano P.;
RT "Mena, a relative of VASP and Drosophila Enabled, is implicated in the
RL control of microfilament dynamics.";
Cell 87:227-239(1996).
CC -!- FUNCTION: MAY BE INVOLVED IN MICROFILAMENT ASSEMBLY AND CELL
CC MOTILITY. INDUCES THE FORMATION OF F-ACTIN RICH OUTGROWTHS IN
CC FIBROBLASTS. BINDS PROFILIN.
CC -!- SUBCELLULAR LOCATION: LOCALIZED TO FOCAL ADHESIONS AND, TO A
CC LESSER EXTENT, LEADING EDGES AND STRESS FIBERS.
CC -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS; MENA, MENA+, MENA++ AND MENA+++
CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
DR EMBL; U72520; AAC52863.1; -
DR EMBL; U72521; AAC52864.1; -
DR EMBL; U72522; AAC52865.1; -
DR EMBL; U72523; AAC52866.1; -
DR MGD; MGI:108360; Enah.
DR InterPro; IPR000697; EVH1.
DR InterPro; IPR000156; RanBP1.
DR Pfam; PF00568; WH1; 1.
DR SMART; SM00160; RanBD; 1.
DR SMART; SM00461; WH1; 1.
KW Alternative splicing; Phosphorylation.
FT DOMAIN 442 464 POLY-PRO.
FT DOMAIN 542 552 POLY-PRO.
FT DOMAIN 562 574 POLY-PRO.
FT DOMAIN 578 589 POLY-PRO.
FT DOMAIN 593 605 POLY-PRO.
FT DOMAIN 686 689 POLY-ARG.
FT MOD_RES 255 255 PHOSPHORYLATION (BY CAPK AND CGPK) (BY
FT SIMILARITY).
FT MOD_RES 637 637 PHOSPHORYLATION (BY CAPK AND CGPK) (BY
FT VARSPLIC 117 131 SIMILARITY).
FT VARSPLIC 117 135 MISSING (IN ISOFORM MENA++).
FT VARSPLIC 132 135 MISSING (IN ISOFORM MENA AND ISOFORM
FT VARSPLIC 259 500 CIPC -> VFYL (IN ISOFORM MENA++).
FT SEQUENCE 802 AA; 85844 MW; 592BB975EE20F7F CRC64;
```

```
Query Match 15.4%; Score 158; DB 11; Length 802;
Best Local Similarity 26.3%; Pred. NO. 0.00079;
Matches 59; Conservative 15; Mismatches 62; Indels 88; Gaps 10;

QY 13 NKTERSTAPLNT-----QISALRNDP----- 33
Db 348 NKNSRSPSVNTFPSSQPPAAKSCAWPTSNFSLPSPFPMISSPPGKATGPRVLPVCVS 407
QY 34 KPLPOQPPAPANODQN-----SSQNTRLQPTPIIPAPAKPAAPRPL----- 76
Db 408 SPVPQMPSPPTAPNGSLDSVTYPVSPPTSGPAARPPPPPPPPPPPPPLPPPLPLAS 467
QY 77 -----DRESGVENKLI-----PSVGSPPASS-TPLPPD-GTGPNSTNNRAV- 116
Db 468 LSHCGSQASPPPGTGLASTPSSKPSVLPSPSAGAPASAEPLNPELGDSSASEPGIQAAS 527
QY 117 ----TPVSGSSNSSADPKAPPPPVSSGE-----PPTLGENP 150
Db 528 QPAESPTPOG--LVLGPPAPPPPPPLPSPGATAYASALPPPPGPPP 569
```

Search completed: October 15, 2003, 10:35:50
Job time : 115.475 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 21.5175 Seconds
(without alignments)
258.182 Million cell updates/sec

Title: US-09-915-543-15_COPY_349_383

Perfect score: 178

Sequence: 1 DGLSQEQLHRRSLQTLRDIQRLFPDKEFTGA 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

```
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	178	100.0	1426	23 AAB71229	Human legless homo
2	178	100.0	1435	22 ABB11808	Human BCL9 homolog
3	109	61.2	320	23 AAU78461	Mouse beta-catenin
4	109	61.2	1494	23 AAU78460	Mouse beta-catenin
5	107	60.1	738	23 AAU78463	Human beta-catenin
6	107	60.1	1115	23 AAB71230	Human legless homo
7	66	37.1	1429	22 ABB58779	Drosophila melanog
8	66	37.1	1464	23 AAB71228	D. melanogaster lg
9	57.5	32.3	603	24 AAB725853	Aspergillus fumiga

10	57.5	32.3	618	24 AAB726453	Aspergillus fumiga
11	56	31.5	1294	22 ABB63502	Drosophila melanog
12	55	30.9	294	22 AAB95073	Human protein sequ
13	55	30.9	390	23 AAG96285	Human ovarian canc
14	55	30.9	433	22 AAG63851	Amino acid sequenc
15	55	30.9	433	22 AAG63852	Amino acid sequenc
16	55	30.9	433	22 AAB68522	Human GTP-binding
17	55	30.9	464	22 AAB92828	Human protein sequ
18	55	30.9	464	23 AAG96281	Human ovarian canc
19	55	30.9	643	23 AAG96282	Human ovarian canc
20	55	30.9	718	23 AAG96284	Human ovarian canc
21	55	30.9	751	23 AAG96283	Human ovarian canc
22	54.5	30.6	174	22 AAB86346	Murine bHLH transcr
23	54.5	30.6	174	22 AAB86348	Murine bHLH transcr
24	53	29.8	819	22 AAU69744	Thermus thermophil
25	52.5	29.5	181	22 AAB86345	Human bHLH transcr
26	52.5	29.5	181	22 AAB86347	Human bHLH transcr
27	52	29.2	959	21 AAY53051	Human secreted pro
28	51.5	28.9	210	23 ABB43969	Adenomatous polypo
29	51	28.7	329	21 AAG13171	Arabidopsis thalia
30	51	28.7	434	22 AAU33491	Enterococcus faeca
31	51	28.7	440	23 ABB47770	Enterococcus faeca
32	51	28.7	448	22 AAU35058	Protein #16 relate
33	51	28.7	500	22 AAG89346	Enterococcus faeca
34	51	28.7	1464	22 ABB71111	Drosophila melanog
35	51	28.7	1755	20 AAY41139	Mouse mammary tumo
36	50.5	28.4	96	22 ABB65662	Drosophila melanog
37	50.5	28.4	479	23 ABB43965	Drosophila melanog
38	50.5	28.4	675	21 AAY54052	Unidentified prote
39	50.5	28.4	675	21 AAY54053	An angiogenesis-as
40	50.5	28.4	1183	22 ABB58769	A variant of an an
41	50	28.1	187	21 AAB38555	Drosophila melanog
42	50	28.1	221	21 AAB07850	Human secreted pro
43	50	28.1	223	22 AAB62225	Amino acid sequenc
44	50	28.1	275	21 AAB24369	Glycine max glutat
45	50	28.1	342	20 AAY42781	Bacillus subtilis

ALIGNMENTS

RESULT 1

AAAB71229	ID	AAAB71229 standard; Protein; 1426 AA.
XX	AC	AAAB71229;
XX	DT	18-NOV-2002 (first entry)
XX	DE	Human legless homologue lgs/bcl9 protein.
XX	DE	Legless; human; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;
KW	KW	tissue proliferation; tumour; cytosolic; cellular disorder; colon;
KW	KW	blood disorder; cancer; breast; head and neck cancer; brain; thyroid;
KW	KW	medulloblastoma; skin cancer; tissue regeneration; tissue repair.
XX	OS	Homo sapiens.
XX	PN	US2002086986-A1.
XX	PD	04-JUL-2002.
XX	PF	27-JUL-2001; 2001US-0915543.
XX	PR	28-JUL-2000; 2000US-221502P.
XX	PA	(BASL/) BASLER K.
PA	PA	(BRUN/) BRUNNER E.
PA	PA	(FRCE/) FROESCH B.
PA	PA	(KRAM/) KRAMPES T.
PA	PA	(PETE/) PETER O.
XX	PI	Basler K, Brunner E, Froesch B, Kramps T, Peter O;

XX WPI; 2002-635689/68.
 DR N-PSDB; AAF88467.
 XX
 PT Novel polypeptide useful in therapeutic method for treating disorders
 PT of cell fate such as cell differentiation or cell proliferation -
 XX
 XX Example II; Fig 8B; 4lpp; English.
 PS
 XX This invention describes a novel polypeptide sharing one or more
 CC homologous amino acid domains with the legless (lgs) protein, a
 CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway
 CC involved in the formation and maintenance of spatial arrangements
 CC and proliferation of tissues during development, and in the formation
 CC and growth of many human tumours. The products of the invention have
 CC cytotatic activity and can be used to treat cellular disorders, blood
 CC disorders and cancers caused by over-stimulation of the Wnt pathway,
 CC where the cancerous condition is colon, breast, head and neck, brain,
 CC thyroid, medulloblastoma or skin cancer. The product could also be used
 CC to promote tissue regeneration and repair. This sequence represents the
 CC human legless (lgs) protein homologue lgs/bcl9 described in the
 CC disclosure of the invention.
 XX
 SQ Sequence 1426 AA;
 Query Match 100.0%; Score 178; DB 23; Length 1426;
 Best Local Similarity 100.0%; Pred. No. 8.6e-16;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DGLSQEQLHRSLSQTLRDQIQLMFLPDEKEFTGA 35
 Db 349 DGLSQEQLHRSLSQTLRDQIQLMFLPDEKEFTGA 383
 RESULT 2
 ABB11808
 ID ABB11808 standard; peptide; 1435 AA.
 XX
 AC ABB11808;
 XX
 DT 11-JAN-2002 (first entry)
 XX
 DE Human BCL9 homologue, SEQ ID NO:2178.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
 KW chronic inflammatory condition; proliferative retinopathy;
 KW atherosclerosis; coronary heart disease; arterial ischaemia;
 KW bone disorder; osteoporosis; vascular growth disorder;
 KW tissue regeneration; wound healing; infection; immune disorder;
 KW cell culture; drug screening; gene therapy; antiinflammatory;
 KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
 KW cytotatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
 KW antifungal; vulnery; antiulcer.
 XX
 OS Homo sapiens.
 XX
 PN WO200157188-A2.
 XX
 PD 09-AUG-2001.
 XX
 XX 05-FEB-2001; 2001WO-US03800.
 XX
 XX 03-FEB-2000; 2000US-0496914.
 PR
 PR 27-APR-2000; 2000US-0560875.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Drmanac RT;
 PI
 XX

DR WPI; 2001-457740/49.
 DR N-PSDB; ABA09052.
 XX
 PT Human proteins and DNA encoding sequences useful for preventing,
 PT treating or ameliorating a medical condition in a mammalian subject
 PT e.g. arthritis and cancer -
 XX
 XX Claim 20; Page 256-257; 1963pp; English.
 PS
 XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
 CC invention also relates to vectors and recombinant host cells comprising a
 CC nucleotide of the invention, methods of producing the novel polypeptides,
 CC antibodies against the polypeptides, methods of detecting the nucleotides,
 CC or polypeptides in a sample, and methods of identifying compounds which
 CC bind to polypeptides of the invention. Although novel, many of the
 CC polypeptides of the invention have homology to known proteins, thereby
 CC giving an insight into their probable biological activities, and hence
 CC potential therapeutic applications. The polypeptides of the invention may
 CC have various activities, including cytokine, cell proliferation or cell
 CC differentiation activities; stem cell growth factor activity;
 CC haematopoiesis regulatory activity; tissue growth activity;
 CC immunomodulatory activity; activin- or inhibin-related activities;
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
 CC thrombolytic activities; receptor or ligand activities; or may be
 CC involved in oncogenesis, cancer cell proliferation or metastasis.
 CC Depending on their biological activities, polypeptides and nucleotides of
 CC the invention are useful for preventing, treating or ameliorating medical
 CC conditions, e.g., by protein or gene therapy. Such conditions include
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
 CC vascular growth. Polypeptides involved with tissue regeneration and
 CC repair (or nucleic acids encoding them) may be used to promote wound
 CC healing (e.g., of burns, incisions and ulcers), while those with
 CC immunomodulatory activities may be used in the treatment of viral,
 CC bacterial and fungal infections in addition to immune disorders.
 CC Polypeptides with growth factor activity may be used in cell cultures to
 CC promote cell growth. For example, such polypeptides may be used to
 CC manipulate stem cells in culture to give rise to neuroepithelial cells
 CC that can be used to augment or replace cells damaged by illness,
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides
 CC may also be used in the diagnosis of the above conditions, and in drug
 CC screening techniques. The present sequence represents a novel human
 CC polypeptide of the invention.
 XX
 SQ Sequence 1435 AA;
 Query Match 100.0%; Score 178; DB 22; Length 1435;
 Best Local Similarity 100.0%; Pred. No. 8.7e-16;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DGLSQEQLHRSLSQTLRDQIQLMFLPDEKEFTGA 35
 Db 389 DGLSQEQLHRSLSQTLRDQIQLMFLPDEKEFTGA 423
 RESULT 3
 AAU78461
 ID AAU78461 standard; Protein; 320 AA.
 XX
 AC AAU78461;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Mouse beta-catenin nuclear localised protein #2.
 DE
 XX Mouse; beta-catenin nuclear localised protein; cancer;
 KW gene therapy; EST; expressed sequence tag.
 KW
 XX Mus musculus.
 OS
 XX

```

PN WO200224738-A1.
XX
PD 28-MAR-2002.
XX
PF 19-SEP-2001; 2001WO-JP08140.
XX
PR 22-SEP-2000; 2000JP-0287876.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Akiyama T, Adachi S;
XX
DR WPI; 2002-330014/36.
XX
DR N-PSDB; ABK47632.
XX
PT New beta-catenin nuclear localised protein for diagnosis and treatment
PT of diseases associated with nuclear localisation of beta-catenin e.g.
PT cancer
XX
PS Claim 2; Page 91-92; 113pp; Japanese.
XX
CC The invention relates to a beta-catenin nuclear localised protein
CC and DNA encoding the protein. The protein and encoding DNA are
CC applicable in diagnosis and treatment of diseases associated with
CC nuclear localisation of beta-catenin e.g. cancer, including gene
CC therapy. The present sequence represents the amino acid sequence of
CC mouse beta-catenin nuclear localised protein #2.
XX
SQ Sequence 320 AA;

Query Match 61.2%; Score 109; DB 23; Length 320;
Best Local Similarity 84.0%; Pred. No. 6.3e-07;
Matches 21; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 DGLSQEQLHRSLSQTLRDQIORML 25
DB 150 EGLSQEQLHRSLSQTLRDIERLL 174

RESULT 4
AAU78460
ID AAU78460 standard; Protein; 1494 AA.
XX
AC AAU78460;
XX
DT 02-JUL-2002 (first entry)
XX
DE Mouse beta-catenin nuclear localised protein.
XX
KW Mouse; beta-catenin nuclear localised protein; cancer;
KW gene therapy; EST; expressed sequence tag.
XX
OS Mus musculus.
XX
PN WO200224738-A1.
XX
PD 28-MAR-2002.
XX
PF 19-SEP-2001; 2001WO-JP08140.
XX
PR 22-SEP-2000; 2000JP-0287876.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Akiyama T, Adachi S;
XX
DR WPI; 2002-330014/36.
XX
DR N-PSDB; ABK47631.
XX
PT New beta-catenin nuclear localised protein for diagnosis and treatment
PT of diseases associated with nuclear localisation of beta-catenin e.g.
PT cancer
XX

Claim 1; Page 81-88; 113pp; Japanese.
XX
CC The invention relates to a beta-catenin nuclear localised protein
CC and DNA encoding the protein. The protein and encoding DNA are
CC applicable in diagnosis and treatment of diseases associated with
CC nuclear localisation of beta-catenin e.g. cancer, including gene
CC therapy. The present sequence represents the amino acid sequence of
CC mouse beta-catenin nuclear localised protein.
XX
SQ Sequence 1494 AA;

Query Match 61.2%; Score 109; DB 23; Length 1494;
Best Local Similarity 84.0%; Pred. No. 3.5e-06;
Matches 21; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 DGLSQEQLHRSLSQTLRDQIORML 25
DB 394 EGLSQEQLHRSLSQTLRDIERLL 418

RESULT 5
AAU78463
ID AAU78463 standard; Protein; 738 AA.
XX
AC AAU78463;
XX
DT 02-JUL-2002 (first entry)
XX
DE Human beta-catenin nuclear localised protein #2.
XX
KW Mouse; beta-catenin nuclear localised protein; cancer;
KW gene therapy; EST; expressed sequence tag.
XX
OS Homo sapiens.
XX
PN WO200224738-A1.
XX
PD 28-MAR-2002.
XX
PF 19-SEP-2001; 2001WO-JP08140.
XX
PR 22-SEP-2000; 2000JP-0287876.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Akiyama T, Adachi S;
XX
DR WPI; 2002-330014/36.
XX
DR N-PSDB; ABK47638.
XX
PT New beta-catenin nuclear localised protein for diagnosis and treatment
PT of diseases associated with nuclear localisation of beta-catenin e.g.
PT cancer
XX
PS Claim 8; Page 102-105; 113pp; Japanese.
XX
CC The invention relates to a beta-catenin nuclear localised protein
CC and DNA encoding the protein. The protein and encoding DNA are
CC applicable in diagnosis and treatment of diseases associated with
CC nuclear localisation of beta-catenin e.g. cancer, including gene
CC therapy. The present sequence represents the amino acid sequence of
CC human beta-catenin nuclear localised protein #2.
XX
SQ Sequence 738 AA;

Query Match 60.1%; Score 107; DB 23; Length 738;
Best Local Similarity 87.5%; Pred. No. 3e-06;
Matches 21; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 GLSQEQLHRSLSQTLRDQIORML 25
DB 1 GLSQEQLHRSLSQTLRDIERLL 24

```

```

RESULT 6
AAB71230
ID AAB71230 standard; Protein; 1115 AA.
XX
AC AAB71230;
XX
18-NOV-2002 (first entry)
XX
Human legless homologue higs-1 partial protein.
XX
Legless; human; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;
KW tissue proliferation; tumour; cytosstatic; cellular disorder; colon;
KW blood disorder; cancer; breast; head and neck cancer; brain; thyroid;
KW medulloblastoma; skin cancer; tissue regeneration; tissue repair.
XX
Homo sapiens.
OS
US2002086986-A1.
PN
04-JUL-2002.
PD
27-JUL-2001; 2001US-0915543.
PF
28-JUL-2000; 2000US-221502P.
PR
(BASL/) BASLER K.
PA (BRUN/) BRUNNER E.
PA (FROE/) FROESCH B.
PA (KRAM/) KRAMPS T.
PA (PETE/) PETER O.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
WPI; 2002-635689/68.
DR N-PSDB; AAF88468.
XX
Novel polypeptide useful in therapeutic method for treating disorders
PT of cell fate such as cell differentiation or cell proliferation -
PT
XX
Example II; Fig 10B; 41pp; English.
XX
This invention describes a novel polypeptide sharing one or more
CC homologous amino acid domains with the legless (lgs) protein, a
CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway
CC involved in the formation and maintenance of spatial arrangements
CC and proliferation of tissues during development, and in the formation
CC and growth of many human tumours. The products of the invention have
CC cytosstatic activity and can be used to treat cellular disorders, blood
CC disorders and cancers caused by over-stimulation of the Wnt pathway,
CC where the cancerous condition is colon, breast, head and neck, brain,
CC thyroid, medulloblastoma or skin cancer. The product could also be used
CC to promote tissue regeneration and repair. This sequence represents the
CC human legless (lgs) protein homologue higs-1 described in the
CC disclosure of the invention.
XX
SQ Sequence 1115 AA;
Query Match 60.1%; Score 107; DB 23; Length 1115;
Best Local Similarity 87.5%; Pred. No. 4.8e-06;
Matches 21; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
OY 2 GLSKEQLHRRSLQTLRDIQRL 25
DB 76 GLSKEQLHRRSLQTLRDIQRL 99
XXXXXXXXXXXXXXXXXXXXXXXXXXXX
RESULT 7
ABB58779
ID ABB58779 standard; Protein; 1429 AA.
XX
AC ABB58779;
XX
26-MAR-2002 (first entry)
XX
Drosophila melanogaster polypeptide SEQ ID NO 3129.
XX
Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
Drosophila melanogaster.
OS
WO200171042-A2.
PN
27-SEP-2001.
PD
23-MAR-2001; 2001WO-US09231.
PF
23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
(PETK ) PE CORP NY.
PA
Venter JC, Adams M, Li PWD, Myers EW;
XX
WPI; 2001-656860/75.
DR N-PSDB; ABL02882.
XX
New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
Disclosure; SEQ ID NO 3129; 21pp + Sequence Listing; English.
XX
The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
XX
The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
XX
SQ Sequence 1429 AA;
Query Match 37.1%; Score 66; DB 22; Length 1429;
Best Local Similarity 31.4%; Pred. No. 3.1;
Matches 11; Conservative 10; Mismatches 14; Indels 0; Gaps 0;
OY 1 DGLSQEQLHRRSLQTLRDIQRLPFDEKEFTGA 35
DB 520 ENLTPOQRHREQLAKIKKMNQFLFPENENSVGA 554
XXXXXXXXXXXXXXXXXXXXXXXXXXXX
RESULT 8
AAB71228
ID AAB71228 standard; Protein; 1464 AA.
XX
AC AAB71228;
XX
18-NOV-2002 (first entry)
XX
D. melanogaster lgs protein.
DE
XX
Legless; fruitfly; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;
KW tissue proliferation; tumour; cytosstatic; cellular disorder; colon;
KW blood disorder; cancer; breast; head and neck cancer; brain; thyroid;
KW medulloblastoma; skin cancer; tissue regeneration; tissue repair.
XX
Drosophila melanogaster.
OS
US2002086986-A1.
PN

```

PD 04-JUL-2002.

XX 27-JUL-2001; 2001US-0915543.

PF 28-JUL-2000; 2000US-221502P.

XX (BASL/) BASLER K.

PA (BRUN/) BRUNNER E.

PA (FROE/) FROESCH B.

PA (KRAM/) KRAMPS T.

PA (PETE/) PETER O.

XX Basler K, Brunner E, Froesch B, Kramps T, Peter O;

PI WPI; 2002-535689/68.

XX N-PSOB; AAF88466.

DR Novel polypeptide useful in therapeutic method for treating disorders

XX of cell fate such as cell differentiation or cell proliferation -

XX Example II; Fig 2; 41pp; English.

XX This invention describes a novel polypeptide sharing one or more

CC homologous amino acid domains with the legless (lgs) protein, a

CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway

CC involved in the formation and maintenance of spatial arrangements

CC and proliferation of tissues during development, and in the formation

CC and growth of many human tumors. The products of the invention have

CC cytoskeletal activity and can be used to treat cellular disorders, blood

CC disorders and cancers caused by over-stimulation of the Wnt pathway,

CC where the cancerous condition is colon, breast, head and neck, brain,

CC thyroid, medulloblastoma or skin cancer. The product could also be used

CC to promote tissue regeneration and repair. This sequence represents the

CC Drosophila melanogaster (fruitfly) legless (lgs) protein described in

CC the disclosure of the invention.

XX Sequence 1464 AA;

SQ Query Match 37.1%; Score 66; DB 23; Length 1464;

Best Local Similarity 31.4%; Pred. No. 3.2;

Matches 11; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 1 DGLSQQLHRERSLQTLRDIQRLMFLPDEKEFTGA 35

DB 515 ENLTQQQRHREQLAKIKKMNQFLPENNSVGA 549

RESULT 9

ABJ25853

XX ABJ25853 standard; Protein; 603 AA.

AC ABJ25853;

XX 16-APR-2003 (first entry)

DE Aspergillus fumigatus essential gene protein #511.

XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;

KW cancer; contamination; biofilm; antibody; immune response.

XX Aspergillus fumigatus.

OS Aspergillus fumigatus.

XX WO200286090-A2.

PN 31-OCT-2002.

XX 23-APR-2002; 2002WO-US13142.

PF 23-APR-2001; 2001US-285697P.

PR 27-APR-2001; 2001US-287066P.

PR 05-JUN-2001; 2001US-295890P.

PR 09-JUL-2001; 2001US-303899P.

PR 31-AUG-2001; 2001US-316362P.

XX (ELIT-) ELITRA PHARM INC.

XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;

XX WPI; 2003-093124/08.

XX New purified or isolated nucleic acids of essential genes of

PT Aspergillus fumigatus, useful for treating or preventing infections by

PT A. fumigatus, or for treating a non-infectious disease in a subject

PT e.g. cancer -

XX Disclosure; Page -; 175pp; English.

XX The invention relates to novel purified or isolated nucleic acids of

CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of

CC the invention are used to treat or prevent infections by a pathogenic

CC organism such as A. fumigatus, to treat a non-infectious disease in a

CC subject (e.g. cancer), to prevent or contain contamination of an object

CC by A. fumigatus, or to prevent or inhibit formation on a surface of a

CC biofilm comprising A. fumigatus. The polynucleotides are useful for

CC expressing recombinant protein for characterisation, screening or

CC therapeutic use, as markers for host tissues in which the pathogenic

CC organisms invade or reside, for comparing with the DNA sequence of A.

CC fumigatus to identify duplicated genes or paralogues having the same or

CC similar biochemical activity and/or function, for comparing with DNA

CC sequences of other related or distant pathogenic organisms to identify

CC potential orthologous essential or virulence genes, for selecting and

CC making oligomers for attachment to a nucleic acid array for examination

CC of expression patterns, for raising anti-protein antibodies, as an

CC antigen to raise anti-DNA antibodies or to elicit another immune

CC response, and for identifying polynucleotides encoding the other protein

CC with which binding occurs or to identify inhibitors of the binding

CC interaction. The polypeptides may be used to raise antibodies or to

CC elicit immune response, as a reagent in assays designed to quantitatively

CC determine levels of the protein in biological fluids, as a marker for

CC host tissues in which pathogenic organism invade or reside, and to

CC isolate correlative receptors or ligands in the case or virulence

CC factors. This sequence represents a protein of one of the essential genes

CC of Aspergillus fumigatus of the invention.

XX Sequence 603 AA;

SQ Query Match 32.3%; Score 57.5; DB 24; Length 603;

Best Local Similarity 31.4%; Pred. No. 18;

Matches 11; Conservative 11; Mismatches 10; Indels 3; Gaps 1;

QY 1 DGLSQQLHRERSLQTLRDIQRLMFLPDEKEFTGA 35

DB 45 DGVETEKIREKD---EVEKLERMLFGDGEFVGA 76

RESULT 10

ABJ26453

XX ABJ26453 standard; Protein; 618 AA.

AC ABJ26453;

XX 16-APR-2003 (first entry)

DE Aspergillus fumigatus essential gene protein #1111.

XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;

KW cancer; contamination; biofilm; antibody; immune response.

XX Aspergillus fumigatus.

OS WO200286090-A2.

PN 31-OCT-2002.

XX 23-APR-2002; 2002WO-US13142.

[illegible]

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 16943; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 294 AA;
 Query Match 30.9%; Score 55; DB 22; Length 294;
 Best Local Similarity 52.2%; Pred. No. 18;
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
 QY 2 GLSQQLHREHRSIQTLRDQRM 24
 || || | :||:|:
 Db 224 GLRTEGLFRRSASVQTVREIQRL 246
 RESULT 13
 ABG96285
 ID ABG96285 standard; Protein; 390 AA.
 XX
 AC ABG96285;
 XX
 DT 11-DEC-2002 (first entry)
 XX
 DE Human ovarian cancer marker M355.
 XX
 KW Human; ovarian cancer; marker; cancer; familial history; brain disorder;
 KW central nervous system disorder; bacterial meningitis; viral meningitis;
 KW Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;
 KW brain herniation; inflammation; encephalitis; testicular disorder;
 KW nontuberculous granulomatous orchitis; connective tissue disorder;
 KW heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;
 KW histological type; carcinogenic; ovarian cancer marker.
 XX
 OS Homo sapiens.
 XX
 PN WO200271928-A2.
 XX
 PD 19-SEP-2002.
 XX
 PF 14-MAR-2002; 2002WO-US07826.
 XX
 XX 14-MAR-2001; 2001US-276025P.
 PR 14-MAR-2001; 2001US-276026P.
 PR

PR 10-AUG-2001; 2001US-311732P.
 PR 19-SEP-2001; 2001US-323580P.
 PR 26-SEP-2001; 2001US-324967P.
 PR 26-SEP-2001; 2001US-325102P.
 PR 26-SEP-2001; 2001US-325149P.
 XX (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Monahan JE, Gannavarapu M, Hoersch S, Kamatkar S, Kovatis SG;
 PI Meyers RE, Morrissy MP, Olandt PJ, Sen A, Vieby PO, Mills GB;
 PI Bast RC, Lu K, Schmandt RE, Zhao X, Glatt K;
 XX
 WPI: 2002-723277/78.
 N-PSDB; ABS76377.
 DR
 DR
 XX Assessing whether a patient is afflicted with ovarian cancer, useful in
 PT assessing the stage or progression of the disease, comprises comparing
 PT the expression level of a cancer marker in a sample from a patient and
 PT from a non cancer patient -
 XX
 PS Disclosure; Page 134-135; 481pp; English.
 XX
 CC The present invention relates to a new method for assessing whether a
 CC patient is afflicted with ovarian cancer. The method involves comparing
 CC the expression level of a marker in a patient sample and the normal level
 CC of expression of the marker in a control non-ovarian cancer sample, where
 CC the marker is selected from 363 cancer markers described in the
 CC specification. The method of the invention is useful in diagnosing or
 CC characterising cancer, in detecting the presence of cancer as early as
 CC possible, and the recurrence of ovarian cancer. The method may also be of
 CC particular use with patients having an enhanced risk of developing
 CC ovarian cancer (e.g. patients having a familial history of ovarian
 CC cancer). The cancer markers may be used in the management and treatment
 CC of e.g. brain and central nervous system disorders (e.g. bacterial and
 CC viral meningitis, Alzheimer's disease or Parkinson's disease), brain
 CC disorders (e.g. cerebral oedema, hydrocephalus or brain herniations),
 CC inflammations (e.g. bacterial or viral meningitis or encephalitis),
 CC testicular disorders (e.g. nontuberculous granulomatous orchitis),
 CC connective tissue disorders, or heart disorders (e.g. ischaemic heart
 CC disease or atherosclerosis). The compositions and methods may also be
 CC used in assessing the histological type of neoplasm associated with
 CC ovarian cancer, monitoring the progression of ovarian cancer,
 CC determining whether ovarian cancer has metastasized or is likely to
 CC metastasize, selecting a composition for inhibiting ovarian cancer,
 CC assessing the ovarian carcinogenic potential of a compound, or
 CC inhibiting ovarian cancer or at risk of developing ovarian cancer. The
 CC present amino acid sequence represents one of the ovarian cancer markers
 CC described in the invention.
 XX
 SQ Sequence 390 AA;
 Query Match 30.9%; Score 55; DB 23; Length 390;
 Best Local Similarity 52.2%; Pred. No. 25;
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
 QY 2 GLSQQLHREHRSIQTLRDQRM 24
 || || | :||:|:
 Db 181 GLRTEGLFRRSASVQTVREIQRL 203
 RESULT 14
 AAG63851
 ID AAG63851 standard; Protein; 433 AA.
 XX
 AC AAG63851;
 XX
 DT 29-OCT-2001 (first entry)
 XX
 DE Amino acid sequence of human GTPase activating protein GTPAP1.
 XX
 KW GTPase activating protein; GTPAP1; cell signalling; immune disorder;
 KW cell proliferative disorder; cancer; colon cancer; arteriosclerosis;
 KW diabetes; psoriasis; hepatitis; multiple sclerosis; gene therapy.
 KW

XX Homo sapiens. Location/Qualifiers
 XX Key Modified-site 60 /note= "potential protein kinase phosphorylation site for tyrosine kinase"
 PH Modified-site 101 /note= "potential protein kinase phosphorylation site for tyrosine kinase"
 FT Modified-site 129 /note= "potential protein kinase phosphorylation site for protein kinase C"
 FT Modified-site 169 /note= "potential protein kinase phosphorylation site for casein kinase II"
 FT Domain 210..235 /note= "GTPase-activator domain"
 FT Domain 310..350 /note= "GTPase-activator domain"
 FT Modified-site 239 /note= "potential protein kinase phosphorylation site for casein kinase II"
 FT Modified-site 239 /note= "potential protein kinase phosphorylation site for protein kinase C"
 FT Modified-site 292 /note= "potential protein kinase phosphorylation site for casein kinase II"
 FT Modified-site 297 /note= "potential protein kinase phosphorylation site for protein kinase C"
 FT Modified-site 309 /note= "potential protein kinase phosphorylation site for casein kinase II"
 FT Domain 310..350 /note= "GTPase-activator domain"
 FT Modified-site 315 /note= "potential protein kinase phosphorylation site for tyrosine kinase"
 FT Modified-site 338 /note= "potential N-linked glycosylation site"
 FT Modified-site 382 /note= "potential protein kinase phosphorylation site for casein kinase II"
 XX WO200161010-A2.
 PN 23-AUG-2001.
 XX 15-FEB-2001; 2001WO-US05075.
 XX 18-FEB-2000; 2000US-0507765.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Klingler TM, Stewart EA, Yue H, Baughn MR;
 XX WPI; 2001-522598/57.
 DR N-PSDB; AAH74978.
 XX A substantially purified GTPase activating protein useful for treating or preventing cell signalling, immune and cell proliferative disorders, including cancer especially colon cancer -
 XX Claim 14; Fig 1A-E; 78pp; English.
 XX The present sequence represents a human GTPase activating protein, designated GTPAP1. GTPAP polynucleotides and polypeptides are useful for treating or preventing a disease or condition associated with altered expression of GTPase activating proteins, especially cell signalling, immune and cell proliferative disorders, including cancer especially colon cancer. Examples of diseases treatable include

CC arteriosclerosis, diabetes, psoriasis, hepatitis and multiple sclerosis.
 CC GTPAP1 polynucleotides are also useful for gene therapy treatments of the diseases.
 XX Sequence 433 AA;
 SQ Query Match 30.9%; Score 55; DB 22; Length 433;
 Best Local Similarity 52.2%; Pred. NO. 28;
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
 QY 2 GLSQEQLEHRSLSQTLRDQRM 24
 Db 224 GLRTEGLFRKSASVQTVREIQRL 246
 RESULT 15
 AAG63852
 ID AAG63852 standard; Protein; 433 AA.
 XX AAG63852;
 AC 29-OCT-2001 (first entry)
 XX Amino acid sequence of human GTPase activating protein GTPAP2.
 DE GTPase activating protein; GTPAP2; cell signalling; immune disorder;
 KW cell proliferative disorder; cancer; colon cancer; arteriosclerosis;
 XX diabetes; psoriasis; hepatitis; multiple sclerosis; gene therapy.
 OS Homo sapiens.
 XX Key Location/Qualifiers
 FT Modified-site 60 /note= "potential protein kinase phosphorylation site for tyrosine kinase"
 FT Modified-site 101 /note= "potential protein kinase phosphorylation site for tyrosine kinase"
 FT Modified-site 129 /note= "potential protein kinase phosphorylation site for protein kinase C"
 FT Modified-site 169 /note= "potential protein kinase phosphorylation site for casein kinase II"
 FT Domain 210..235 /note= "GTPase-activator domain"
 FT Domain 310..350 /note= "GTPase-activator domain"
 FT Modified-site 239 /note= "potential protein kinase phosphorylation site for casein kinase II"
 FT Modified-site 239 /note= "potential protein kinase phosphorylation site for protein kinase C"
 FT Modified-site 292 /note= "potential protein kinase phosphorylation site for casein kinase II"
 FT Modified-site 297 /note= "potential protein kinase phosphorylation site for protein kinase C"
 FT Modified-site 309 /note= "potential protein kinase phosphorylation site for casein kinase II"
 FT Domain 310..350 /note= "GTPase-activator domain"
 FT Modified-site 315 /note= "potential protein kinase phosphorylation site for tyrosine kinase"
 FT Modified-site 338 /note= "potential N-linked glycosylation site"
 FT Modified-site 382 /note= "potential protein kinase phosphorylation site for casein kinase II"

Search completed: October 15, 2003, 10:30:24
Job time : 23.5175 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 7.08171 Seconds
(without alignments)
209.113 Million cell updates/sec

Title: US-09-915-543-15_COPY_349_383

Perfect score: 178
Sequence: 1 DGLSQEHLERSLQTLRDIQRLFPDEKEFTGA 35

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA: *
1: /cgn2.6/ptodata/1/1aa/5A_COMB.pep: *
2: /cgn2.6/ptodata/1/1aa/5B_COMB.pep: *
3: /cgn2.6/ptodata/1/1aa/6A_COMB.pep: *
4: /cgn2.6/ptodata/1/1aa/6B_COMB.pep: *
5: /cgn2.6/ptodata/1/1aa/pctUS_COMB.pep: *
6: /cgn2.6/ptodata/1/1aa/pctfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	30.9	333	4 US-09-507-765-32	Sequence 32, Appl
2	55	30.9	433	4 US-09-507-765-30	Sequence 30, Appl
3	55	30.9	433	4 US-09-507-765-31	Sequence 31, Appl
4	54.5	30.6	1242	4 US-09-252-991A-25843	Sequence 25843, A
5	53	29.8	819	4 US-09-651-656-15	Sequence 15, Appl
6	53	29.8	819	4 US-09-650-855-15	Sequence 15, Appl
7	52	29.2	578	4 US-09-252-991A-23906	Sequence 23906, A
8	51	28.7	589	4 US-09-252-991A-24834	Sequence 24834, A
9	50.5	28.4	354	4 US-09-252-991A-28779	Sequence 28779, A
10	50	28.1	221	3 US-09-247-373B-54	Sequence 54, Appl
11	50	28.1	275	4 US-09-542-749A-2	Sequence 2, Appl
12	50	28.1	1201	4 US-09-252-991A-32259	Sequence 32259, A
13	49	27.5	524	4 US-09-186-276B-56	Sequence 56, Appl
14	49	27.5	524	4 US-08-842-443-56	Sequence 56, Appl
15	49	27.5	524	4 US-09-186-188B-56	Sequence 56, Appl
16	49	27.5	1027	4 US-09-252-991A-23210	Sequence 23210, A
17	48.5	27.2	1872	1 US-08-188-582-14	Sequence 14, Appl
18	48.5	27.2	1872	1 US-08-646-715-14	Sequence 14, Appl
19	48.5	27.2	1893	1 US-08-188-582-11	Sequence 11, Appl
20	48.5	27.2	1893	1 US-08-646-715-11	Sequence 11, Appl
21	48	27.0	266	4 US-09-107-532A-5226	Sequence 5226, Ap
22	48	27.0	1037	4 US-09-252-991A-22579	Sequence 22579, A
23	48	27.0	1315	4 US-09-252-991A-22746	Sequence 22746, A
24	47.5	26.7	646	4 US-09-252-991A-29240	Sequence 29240, A
25	47	26.4	1000	4 US-09-252-991A-31361	Sequence 31361, A
26	47	26.4	10182	4 US-09-134-001C-3159	Sequence 3159, Ap
27	46.5	26.1	65	1 US-08-227-536-3	Sequence 3, Appl

28 46.5 26.1 65 5 PCT-US95-04682-3 Sequence 3, Appli
29 46.5 26.1 885 4 US-09-107-532A-5104 Sequence 5104, Ap
30 46.5 26.1 935 4 US-09-107-532A-3753 Sequence 3753, Ap
31 46 25.8 178 4 US-09-252-991A-18906 Sequence 18906, A
32 46 25.8 180 4 US-09-252-991A-31012 Sequence 31012, A
33 46 25.8 452 4 US-09-252-991A-23336 Sequence 23336, A
34 46 25.8 485 4 US-09-328-352-6561 Sequence 6561, Ap
35 46 25.8 558 4 US-09-252-991A-18908 Sequence 18908, A
36 46 25.8 2101 1 US-08-466-390-4 Sequence 4, Appli
37 46 25.8 2101 1 US-08-470-950-4 Sequence 4, Appli
38 46 25.8 2101 1 US-08-467-781-4 Sequence 4, Appli
39 46 25.8 2101 1 US-08-195-487-4 Sequence 4, Appli
40 46 25.8 2101 2 US-08-483-924-4 Sequence 4, Appli
41 46 25.8 2101 3 US-09-452-294-1 Sequence 1, Appli
42 46 25.8 2101 5 PCT-US93-06160-4 Sequence 4, Appli
43 45.5 25.6 118 4 US-08-858-207A-395 Sequence 395, App
44 45.5 25.6 832 4 US-09-252-991A-19252 Sequence 19252, A
45 45 25.3 52 2 US-08-690-011A-34 Sequence 34, Appl

ALIGNMENTS

RESULT 1
US-09-507-765-32
; Sequence 32, Application US/09507765
; Patent No. 6509155
; GENERAL INFORMATION:
; APPLICANT: Klingner, Tod M.
; APPLICANT: Stewart, Elizabeth A.
; APPLICANT: Yue, Henry
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS
; FILE REFERENCE: PC-0010 US
; CURRENT APPLICATION NUMBER: US/09/507,765
; CURRENT FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 32
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. 6509155 g6572185
US-09-507-765-32

Query Match 30.9%; Score 55; DB 4; Length 333;
Best Local Similarity 52.2%; Pred. No. 6.9;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQEHLERSLQTLRDIQRL 24
DB 124 GLRTEGLFRRSASVQTVREIQRL 146

RESULT 2
US-09-507-765-30
; Sequence 30, Application US/09507765
; Patent No. 6509155
; GENERAL INFORMATION:
; APPLICANT: Klingner, Tod M.
; APPLICANT: Stewart, Elizabeth A.
; APPLICANT: Yue, Henry
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS
; FILE REFERENCE: PC-0010 US
; CURRENT APPLICATION NUMBER: US/09/507,765
; CURRENT FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 30
; LENGTH: 433

```
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. 6509155 3068538CD1
US-09-507-765-30

Query Match      30.9%; Score 55; DB 4; Length 433;
Best Local Similarity 52.2%; Pred. No. 9.1;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY      2  GLSQEQLHRRSLSQTLRDQIRM 24
Db      224 GLRTEGLFRRSASQTVREIQRL 246

RESULT 3
US-09-507-765-31
; Sequence 31, Application US/09507765
; Patent No. 6509155
; GENERAL INFORMATION:
; APPLICANT: Klinger, Todd M.
; APPLICANT: Stewart, Elizabeth A.
; APPLICANT: Yue, Henry
; TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS
; FILE REFERENCE: PC-0010 US
; CURRENT APPLICATION NUMBER: US/09/507,765
; CURRENT FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 31
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. 6509155 404424.5.pseq
US-09-507-765-31

Query Match      30.9%; Score 55; DB 4; Length 433;
Best Local Similarity 52.2%; Pred. No. 9.1;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY      2  GLSQEQLHRRSLSQTLRDQIRM 24
Db      224 GLRTEGLFRRSASQTVREIQRL 246

RESULT 4
US-09-252-991A-25843
; Sequence 25843, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25843
; LENGTH: 1242
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25843

Query Match      30.6%; Score 54.5; DB 4; Length 1242;
Best Local Similarity 35.1%; Pred. No. 33;

Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY      2  GLSQEQLHRRSLSQTLRDQIRM 24
Db      224 GLRTEGLFRRSASQTVREIQRL 246

RESULT 5
US-09-651-656-15
; Sequence 15, Application US/09651656
; Patent No. 6340566
; GENERAL INFORMATION:
; APPLICANT: MCCUTHEN-MALONEY, SANDRA
; APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY
; TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE
; TITLE OF INVENTION: POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,
; FILE REFERENCE: IL-10689
; CURRENT APPLICATION NUMBER: US/09/651,656
; CURRENT FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/192,764
; PRIOR FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 15
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Thermus thermophilus
US-09-651-656-15

Query Match      29.8%; Score 53; DB 4; Length 819;
Best Local Similarity 42.3%; Pred. No. 34;
Matches 11; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY      6  EQLEHRRSLSQTLRDQIRMFFPDEKE 31
Db      466 EKVPQYRPVQTLKDKQRYTLPEMKE 491

RESULT 6
US-09-650-855-15
; Sequence 15, Application US/09650855
; Patent No. 6365355
; GENERAL INFORMATION:
; APPLICANT: MCCUTHEN-MALONEY, SANDRA
; APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY
; TITLE OF INVENTION: CHIMERIC PROTEINS FOR DETECTION AND QUANTITATION OF DNA
; TITLE OF INVENTION: MUTATIONS, DNA SEQUENCE VARIATIONS, DNA DAMAGE AND DNA
; FILE REFERENCE: IL-10284
; CURRENT APPLICATION NUMBER: US/09/650,855
; CURRENT FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/192,764
; PRIOR FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 15
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Thermus thermophilus
US-09-650-855-15

Query Match      29.8%; Score 53; DB 4; Length 819;
Best Local Similarity 42.3%; Pred. No. 34;
Matches 11; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY      6  EQLEHRRSLSQTLRDQIRMFFPDEKE 31
Db      466 EKVPQYRPVQTLKDKQRYTLPEMKE 491

RESULT 7
US-09-252-991A-23906
```

```
; Sequence 23906, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23906
; LENGTH: 578
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23906

Query Match      29.2%; Score 52; DB 4; Length 578;
Best Local Similarity 84.6%; Pred. No. 32;
Matches 11; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      6 EOLEHRSLSQTL 18
      | | | | | | | | | |
Db      339 ECLEHRSLSQTL 351

RESULT 8
US-09-252-991A-24834
; Sequence 24834, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24834
; LENGTH: 589
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24834

Query Match      28.7%; Score 51; DB 4; Length 589;
Best Local Similarity 37.0%; Pred. No. 44;
Matches 10; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY      5 QEOLHRSLSQTLRDIQRLFPDEKE 31
      : : | | | | | | | | | |
Db      156 EQORNHRREQQLRYRQALRGFGEDAE 182

RESULT 9
US-09-252-991A-28779
; Sequence 28779, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
```

```
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28779
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28779

Query Match      28.4%; Score 50.5; DB 4; Length 354;
Best Local Similarity 40.0%; Pred. No. 30;
Matches 14; Conservative 6; Mismatches 8; Indels 7; Gaps 2;

QY      2 GLSQE-----QLEHRSLSQTLRDIQRLFPDEKE 31
      | | | | | | | | | | | | | | | |
Db      252 GLAQRLLPGOAPHRPRTLRAVRDRRR--PDLRQ 284

RESULT 10
US-09-247-373B-54
; Sequence 54, Application US/09247373B
; Patent No. 6168954
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
; FILE REFERENCE: CL-1108-A
; CURRENT APPLICATION NUMBER: US/09/247,373B
; CURRENT FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 08/924,747
; PRIOR FILING DATE: 1997-09-05
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 54
; LENGTH: 221
; TYPE: PRT
; ORGANISM: SOYBEAN
US-09-247-373B-54

Query Match      28.1%; Score 50; DB 3; Length 221;
Best Local Similarity 36.7%; Pred. No. 21;
Matches 11; Conservative 7; Mismatches 8; Indels 4; Gaps 1;

QY      9 EHRERSLQT----LRDIQRLFPDEKEFTG 34
      | | | | | | | | | | | | | |
Db      116 EEREKSTIEKIWEHLRVVENQCFCGDKKFFG 145

RESULT 11
US-09-542-749A-2
; Sequence 2, Application US/09542749A
; Patent No. 6428981
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Steen T.
; APPLICANT: Christensen, Christina L.
; APPLICANT: Kristensen, Tina
; TITLE OF INVENTION: A Bacillus Protein Production Cell
; FILE REFERENCE: 5861.200-US
; CURRENT APPLICATION NUMBER: US/09/542,749A
; CURRENT FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: 60/130,194
; PRIOR FILING DATE: 1999-04-20
; PRIOR APPLICATION NUMBER: PA 1999 00506
; PRIOR FILING DATE: 1999-04-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-542-749A-2

Query Match      28.1%; Score 50; DB 4; Length 275;
```

Best Local Similarity 37.0%; Pred. No. 26;
Matches 10; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 7 QLEHRSLSQTLRDIQRMFLPDEKFTG 33
Db 91 ELOGKAGMGFLRNQESLFLVSKKNT 117

RESULT 12
US-09-252-991A-32259
; Sequence 32259, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32259
; LENGTH: 1201
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32259

Query Match 28.1%; Score 50; DB 4; Length 1201;
Best Local Similarity 39.1%; Pred. No. 1.3e+02;
Matches 9; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 3 LSQQLHRSLSQTLRDIQRMFL 25
Db 439 VQSQRIHQESLRLQDRRL 461

RESULT 13
US-09-186-276B-56
; Sequence 56, Application US/09186276B
; Patent No. 6388173
; GENERAL INFORMATION:
; APPLICANT: Benfey, Philip
; APPLICANT: DiLaurenzio, Laura
; APPLICANT: Wysocka-Diller, Joanna
; APPLICANT: Malamy, Jocelyn E.
; APPLICANT: Pysch, Leonard
; APPLICANT: Helariutta, Irjo
; TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses Thereof
; FILE REFERENCE: 5914-075-999
; CURRENT APPLICATION NUMBER: US/09/186,276B
; PRIOR FILING DATE: 1998-11-05
; PRIOR APPLICATION NUMBER: 08/842,445
; PRIOR FILING DATE: 1997-04-24
; PRIOR APPLICATION NUMBER: 08/638,617
; PRIOR FILING DATE: 1996-04-26
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 56
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(524)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-186-276B-56

Query Match 27.5%; Score 49; DB 4; Length 524;
Best Local Similarity 30.0%; Pred. No. 73;
Matches 9; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

QY 5 QOQLHRSLSQTLRDIQRMFLPDEKFTG 34
Db 66 QNRVHSENNLSLRELEKQLLDDDDSGG 95

RESULT 14
US-08-842-445-56
; Sequence 56, Application US/08842445A
; Patent No. 6441270
; GENERAL INFORMATION:
; APPLICANT: Benfey et al.
; TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses
; FILE REFERENCE: 5914-056-999
; CURRENT APPLICATION NUMBER: US/08/842,445A
; CURRENT FILING DATE: 1997-04-24
; EARLIER FILING DATE: 08/638,617
; EARLIER FILING DATE: 1996-04-26
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 56
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Plant
US-08-842-445-56

Query Match 27.5%; Score 49; DB 4; Length 524;
Best Local Similarity 30.0%; Pred. No. 73;
Matches 9; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

QY 5 QOQLHRSLSQTLRDIQRMFLPDEKFTG 34
Db 66 QNRVHSENNLSLRELEKQLLDDDDSGG 95

RESULT 15
US-09-186-188B-56
; Sequence 56, Application US/09186188B
; Patent No. 6455672
; GENERAL INFORMATION:
; APPLICANT: Benfey et al.
; TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses
; FILE REFERENCE: 5914-074-999
; CURRENT APPLICATION NUMBER: US/09/186,188B
; CURRENT FILING DATE: 1998-11-05
; PRIOR APPLICATION NUMBER: 08/842,445
; PRIOR FILING DATE: 1997-04-24
; PRIOR APPLICATION NUMBER: 08/638,617
; PRIOR FILING DATE: 1996-04-26
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 56
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Plant
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(524)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-186-188B-56

Query Match 27.5%; Score 49; DB 4; Length 524;
Best Local Similarity 30.0%; Pred. No. 73;
Matches 9; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

QY 5 QOQLHRSLSQTLRDIQRMFLPDEKFTG 34
Db 66 QNRVHSENNLSLRELEKQLLDDDDSGG 95

Search completed: October 15, 2003, 10:32:00

Wed Oct 15 11:40:27 2003

us-09-915-543-15_copy_349_383.ra1

Page 5

Job time : 9.08171 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:32:06 ; Search time 14.8444 Seconds
(without alignments)
379.908 Million cell updates/sec

Title: US-09-915-543-15_COPY_349_383

Perfect score: 178
Sequence: 1 DGLSQLEHRSLSQTLRDQRMFLFPDEKFTGA 35

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 600653 seqs, 161128416 residues

Total number of hits satisfying chosen parameters: 600653

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	178	100.0	35	15	US-10-322-579-5
2	178	100.0	1426	15	US-10-322-579-15
3	107	60.1	1115	15	US-10-322-579-17
4	66	37.1	35	15	US-10-322-579-4
5	57.5	32.3	603	15	US-10-128-714-3511
6	57.5	32.3	618	15	US-10-128-714-8511
7	55	30.9	294	12	US-10-053-248-12
8	55	30.9	333	16	US-10-284-753-32
9	55	30.9	390	15	US-10-097-340-21
10	55	30.9	433	16	US-10-284-753-30
11	55	30.9	433	16	US-10-284-753-31
12	55	30.9	464	15	US-10-037-340-14
13	55	30.9	643	15	US-10-097-340-16
14	55	30.9	718	15	US-10-097-340-19
15	55	30.9	751	15	US-10-097-340-18

```

16 51 28.7 434 9 US-09-815-242-4987
17 51 28.7 448 9 US-09-815-242-10651
18 51 28.7 500 10 US-09-731-872-466
19 51 28.7 500 12 US-09-876-997-466
20 50 28.1 444 10 US-09-738-626-4366
21 50 28.1 1162 9 US-09-815-242-11828
22 49.5 27.8 484 12 US-10-161-051-119
23 49.5 27.8 804 12 US-10-236-055A-28
24 49.5 27.8 2634 15 US-10-227-610-2
25 49 27.5 201 15 US-10-043-487-323
26 49 27.5 461 15 US-10-234-432-24
27 49 27.5 524 15 US-10-253-007-56
28 49 27.5 535 15 US-10-043-487-306
29 49 27.5 893 15 US-10-234-432-86
30 49 27.5 1413 12 US-09-840-743-8
31 49 27.5 2000 13 US-10-010-901-29
32 48.5 27.2 898 15 US-10-043-487-277
33 48 27.0 170 9 US-09-916-790-15
34 48 27.0 182 15 US-10-156-761-14821
35 47.5 26.7 861 9 US-09-815-242-11081
36 47 26.4 529 15 US-10-156-761-10420
37 47 26.4 670 11 US-09-298-523B-63
38 47 26.4 711 11 US-09-298-523B-3
39 47 26.4 810 15 US-10-128-714-8235
40 47 26.4 811 15 US-10-128-714-3235
41 47 26.4 2462 11 US-09-819-104A-5
42 46.5 26.1 100 9 US-09-867-550-2062
43 46 25.8 118 15 US-10-156-761-14264
44 46 25.8 139 14 US-10-062-254-12
45 46 25.8 142 9 US-09-864-761-42510

```

ALIGNMENTS

RESULT 1
US-10-322-579-5
; Sequence 5, Application US/10322579
; Publication No. US20030114413A1

```

; GENERAL INFORMATION:
; APPLICANT: BASLER, Konrad
; APPLICANT: BRUNNER, Erich
; APPLICANT: FROESCH, Barbara
; APPLICANT: KRAMPS, Thomas
; APPLICANT: PETER, Oliver
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATH
; FILE REFERENCE: Q60361
; CURRENT APPLICATION NUMBER: US/10/322,579
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US/09/915,543
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 60/221,502
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Human lgs/bcl9
US-10-322-579-5

```

Query Match 100.0%; Score 178; DB 15; Length 35;
Best Local Similarity 100.0%; Pred. No. 4.6e-17;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGLSQLEHRSLSQTLRDQRMFLFPDEKFTGA 35
|||||
DB 1 DGLSQLEHRSLSQTLRDQRMFLFPDEKFTGA 35
|||||

RESULT 2
US-10-322-579-15

RESOLUTION 4
US-10-322-579-4
: Sequence 4, Application US/10322579

```
QY      1 DGLSQEQLEHRERSLQTLRDIQRMFLFPDEKEFTGA 35
```

Db 45 DGVETEKIREKD---EVEKKLERMLFGDDGEGFVGA 76

RESULT 6

US-10-128-714-8511
; Sequence 8511, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengli
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Broshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10128,714
; PRIOR FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8511
; LENGTH: 618
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8511

Query Match 32.3%; Score 57.5; DB 15; Length 618;
Best Local Similarity 31.4%; Pred. No. 18;
Matches 11; Conservative 11; Mismatches 10; Indels 3; Gaps 1;

QY 1 DGLSQQLERHRSLSQTLRDQIRMLFPDKEFTGA 35

Db 45 DGVETEKIREKD---EVEKKLERMLFGDDGEGFVGA 76

RESULT 7

US-10-053-248-12
; Sequence 12, Application US/10053248
; Publication No. US20030144188A1
; GENERAL INFORMATION:
; APPLICANT: Lip, Biaoyang
; TITLE OF INVENTION: Androgen Regulated Nucleic Acid
; FILE REFERENCE: P-IS 4814
; CURRENT APPLICATION NUMBER: US/10/053,248
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-053-248-12

Query Match 30.9%; Score 55; DB 12; Length 294;
Best Local Similarity 52.2%; Pred. No. 17;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQQLERHRSLSQTLRDQIR 24

Db 224 GLRTEGLFRRSASVQTVREIQRL 246

RESULT 8

US-10-284-753-32
; Sequence 32, Application US/10284753
; Publication No. US20030129655A1
; GENERAL INFORMATION:
; APPLICANT: Klingert, Tod M.
; APPLICANT: Yue, Henry
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS
; FILE REFERENCE: PC-0010-1 CIP
; CURRENT APPLICATION NUMBER: US/10/284,753
; CURRENT FILING DATE: 2002-10-29
; PRIOR APPLICATION NUMBER: 09/507,765
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 32
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20030129655A1 96572185
US-10-284-753-32

Query Match 30.9%; Score 55; DB 16; Length 333;
Best Local Similarity 52.2%; Pred. No. 19;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQQLERHRSLSQTLRDQIR 24

Db 124 GLRTEGLFRRSASVQTVREIQRL 146

RESULT 9

US-10-097-340-21
; Sequence 21, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPURU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; FILE REFERENCE: Assessment, Prevention, and Therapy of Ovarian Cancer
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102

Wed Oct 15 11:40:28 2003

```

; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 751
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(751)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-097-340-18

```

```

Query Match      30.9%; Score 55; DB 15; Length 751;
Best Local Similarity 52.2%; Pred. No. 49;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

```

```

QY      2 GLSQEQLEHREHRSLOTLRDIQRM 24
      |||||  |||||  |||||  |||||
Db      509 GLRTEGLFRRSASVQTVREIQRL 531

```

```

Search completed: October 15, 2003, 10:50:58
Job time : 15.8444 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 8.30739 Seconds
(without alignments)
405.170 Million cell updates/sec

Title: US-09-915-543-15_COPY_349_383
Perfect score: 178
Sequence: 1 DGLSQQLHRRSLQTLRDIQRMFLPDEKFTGA 35

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR,76:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	57	32.0	1034	AB0551	exonuclease SbcC [
2	56	31.5	584	S51882	topoisomerase I-re
3	56	31.5	859	T46372	hypothetical prote
4	55	30.9	287	AE2895	transcription regu
5	55	30.9	295	H97670	hex regulon repres
6	55	30.9	643	B59436	Rho GTPase activat
7	54.5	30.6	1171	S14065	phytochrome B - ri
8	53.5	30.1	4957	T03455	ALR protein - huma
9	53.5	30.1	5262	T03454	ALR protein - huma
10	53	29.8	227	B70438	hypothetical prote
11	53	29.8	376	E69957	gamma-D-glutamyl-L
12	53	29.8	818	S62790	mismatch DNA recog
13	52.5	29.5	242	B70366	hypothetical prote
14	52.5	29.5	705	T24343	hypothetical prote
15	52.5	29.5	1039	T14802	phytochrome B - so
16	52	29.2	332	B47017	probable transcrip
17	52	29.2	332	AD2541	transcription init
18	52	29.2	572	D82984	pyruvate dehydroge
19	52	29.2	1009	S61174	hypothetical prote
20	51	28.7	102	AH0216	conserved hypothet
21	51	28.7	237	A49940	probable alpha hel
22	51	28.7	237	E91056	probable alpha hel
23	51	28.7	237	A85901	hypothetical prote
24	51	28.7	329	D96834	hypothetical prote
25	51	28.7	477	T18801	pol polyprotein -
26	51	28.7	899	GNMYVM	ATP-dependent dsDN
27	51	28.7	1047	G90684	ATP-dependent protei
28	51	28.7	1047	C85535	development protei
29	51	28.7	1464	S58984	

RESULT 1

AB0551 exonuclease SbcC [imported] - Salmonella enterica subsp. enterica serovar Typhi (str. C; Species: Salmonella enterica subsp. enterica serovar Typhi)
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AB0551
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Far, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, A.; Title: Complete genome sequence of a multiple drug resistant Salmonella enterica s.
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AB0551
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1034 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08850.1; PID:gi6501663; GSPDB:GN00176
C:Genetics:
A:Gene: STY0429
C:Superfamily: sbcC protein

ALIGNMENTS

30	50.5	28.4	835	2	AD2441	endopeptidase Clp
31	50	28.1	273	2	H69337	conserved hypothet
32	50	28.1	275	2	H69843	hypothetical prote
33	50	28.1	319	2	S49771	hypothetical prote
34	50	28.1	518	2	G86454	CDS protein F9L11.
35	50	28.1	788	2	S67595	hypothetical prote
36	50	28.1	1162	2	D83454	conserved hypothet
37	50	28.1	1236	2	B36329	hypothetical prote
38	49.5	27.8	266	2	AE1124	1-pyrroline-5-carb
39	49.5	27.8	302	1	TPCHFC	troponin T, cardia
40	49.5	27.8	336	2	S72858	hypothetical prote
41	49.5	27.8	830	2	T18860	hypothetical prote
42	49	27.5	310	2	A84142	L-lactate dehydrog
43	49	27.5	338	2	F81730	glycerol-3-phospha
44	49	27.5	420	2	T41236	conserved hypothet
45	49	27.5	537	2	C90796	hypothetical prote

Query Match 32.0%; Score 57; DB 2; Length 1034;
Best Local Similarity 42.9%; Pred. No. 26;
Matches 12; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 3 LSQQLHRRSLQTLRDIQRMFLPDEK 30
I: |||: | | | | | : | | |
Db 213 LADEQLQQLASLQALTDKRLADQ 240

RESULT 2

S51882 topoisomerase I-related protein TRF4 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein HRC584; protein O0716; protein YOL115W
C:Species: Saccharomyces cerevisiae
C:Date: 03-May-1995 #sequence_revision 03-Aug-1995 #text_change 21-Jul-2000
C:Accession: S51882; S58774; S66811
R:Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
submitted to the EMBL Data Library, January 1995
A:Description: Sequence analysis of a 44kb DNA fragment of yeast chromosome XV inclu-
and a Delta.
A:Reference number: S51848
A:Accession: S51882
A:Molecule type: DNA
A:Residues: 1-584 <VAN>
A:Cross-references: EMBL:Z48149; NID:g663234; PID:g663237
R:Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
yeast 11, 1069-1075, 1995
A:Title: Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV including
a delta element.

A:Reference number: S59156; MUID:96076631; PMID:7502582
A:Accession: S59158
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-584 <VAV>
A:Cross-references: EMBL:Z48149; NID:g663234; PIDN:CAA88145.1; PID:g663237
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1995
R:Sadoiff, B.O.; Heath-Pagliuso, S.; Castano, I.B.; Zhu, Y.; Kieff, F.S.; Christman, M.F.
Genetics 141, 465-479, 1995
A:Title: Isolation of mutants of Saccharomyces cerevisiae requiring DNA topoisomerase I
A:Reference number: S58774; MUID:96109595; PMID:8647385
A:Accession: S58774
A:Molecule type: DNA
A:Residues: 1-584 <SAD>
A:Cross-references: EMBL:U31355; NID:g950225; PIDN:AAC49091.1; PID:g950226
R:Durand, P.; Hilger, F.; Portetelle, D.; Vandenberg, M.
Submitted to the Protein Sequence Database, July 1996
A:Reference number: S66791
A:Accession: S66811
A:Molecule type: DNA
A:Residues: 1-584 <DUR>
A:Cross-references: EMBL:Z74857; NID:g1419986; PID:e251905; PID:g1419987; MIPS:YOL115W
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:TRF4
A:Cross-references: SGD:S0005475; MIPS:YOL115W
A:Map position: 15L
C:Keywords: nucleus

Query Match 31.5%; Score 56; DB 2; Length 584;
Best Local Similarity 36.0%; Pred. No. 19;
Matches 9; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

Qy 4 SEQLEHRSLSQTLRDIQRMLFPD 28
I::I I::I ::I ::I ::I ::I
Db 193 SREEIEIRNOTISTIREAVKQLWPD 217

RESULT 3
T46372
hypothetical protein DKFzP434P181.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46372
R:Ottewaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23031
A:Accession: T46372
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-859 <AAA>
A:Cross-references: EMBL:AL137528
A:Experimental source: adult testis; clone DKFzP434P181
C:Genetics:
A>Note: DKFzP434P181.1

Query Match 31.5%; Score 56; DB 2; Length 859;
Best Local Similarity 33.3%; Pred. No. 29;
Matches 10; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

Qy 5 SEQLEHRSLSQTLRDIQRMLFPDEKEFTG 34
: I I I I : : I I I I
Db 317 ENQRSHQELISQLLSQYMKLLLPDPDEFHG 346

RESULT 4
AE2895
transcription regulator, RpiR family Atu2598 [imported] - Agrobacterium tumefaciens (str
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AE2895
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
Eragle, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan,
transcription regulator, RpiR family Atu2598 [imported] - Agrobacterium tumefaciens (str

Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQEQLEHRSRSLQTLRDQRM 24
 || || | :|||:||||
 Db 434 GLRTEGLFRKSASVQTVREIQR 456

RESULT 7
 S14065
 phytochrome B - rice
 C:Species: Oryza sativa (rice)
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Jul-2000
 C:Accession: S14065
 R:Dehesh, K.; Tepperman, J.; Christensen, A.H.; Quail, P.H.
 Mol. Gen. Genet. 225, 305-313, 1991
 A:Title: phyB is evolutionarily conserved and constitutively expressed in rice seedling
 A:Reference number: S14065; MUID:91172131; PMID:2005872
 A:Accession: S14065
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1171 <DBH>
 A:Cross-references: GB:X57563; NID:g6469490; PIDN:CAA0795.2; PID:g6469491
 C:Genetics:
 A:Gene: phyB
 C:Superfamily: phytochrome; phytochrome homology
 C:Keywords: Chromoprotein; photoreceptor; phytochromobilin; transcription regulation
 F:103-623/Domain: phytochrome homology <PHYT>
 F:364/Binding site: phytochromobilin (Cys) (covalent) #status predicted

Query Match 30.6%; Score 54.5; DB 2; Length 1171;
 Best Local Similarity 48.1%; Pred. No. 63;
 Matches 13; Conservative 4; Mismatches 5; Indels 5; Gaps 1;

QY 3 LSQEQLEHRSRSLQTLRDQRM 29
 :|| :||| :||| :||| :|||
 Db 1019 VSQVMIQLRDLQIRDI-----PDE 1040

RESULT 8
 T03455
 ALR protein - human
 C:Species: Homo sapiens (man)
 C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
 C:Accession: T03455
 R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,
 Oncogene 15, 549-560, 1997
 A:Title: Structure and expression pattern of human ALR, a novel gene with strong homolog
 A:Reference number: Z14954; MUID:97388474; PMID:9247308
 A:Accession: T03455
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: mRNA
 A:Residues: 1-4957 <PRA>
 A:Cross-references: EMBL:AF010404; NID:g2358286; PIDN:AAC51735.1; PID:g2358287
 C:Genetics:
 A:Gene: ALR
 A:Map position: 12
 C:Superfamily: human ALR protein
 C:Keywords: alternative splicing

Query Match 30.1%; Score 53.5; DB 2; Length 4957;
 Best Local Similarity 42.5%; Pred. No. 4.3e+02;
 Matches 17; Conservative 4; Mismatches 12; Indels 7; Gaps 2;

QY 2 GLSQEQLEHRSRSLQTLRD-----IQRLFPDEKEFTCA 35
 |||| :|| :|| :||| :||| :|||
 Db 2091 GLSQEQLEHRSRSLQTLRD-----IQRLFPDEKEFTCA 2129

RESULT 9
 T03454
 ALR protein - human
 C:Species: Homo sapiens (man)
 C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000

C:Accession: T03454
 R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,
 Oncogene 15, 549-560, 1997
 A:Title: Structure and expression pattern of human ALR, a novel gene with strong hom
 A:Reference number: Z14954; MUID:97388474; PMID:9247308
 A:Accession: T03454
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: mRNA
 A:Residues: 1-5262 <PRA>
 A:Cross-references: EMBL:AF010403; NID:g2358284; PIDN:AAC51734.1; PID:g2358285
 C:Genetics:
 A:Gene: ALR
 A:Map position: 12
 C:Superfamily: human ALR protein
 C:Keywords: alternative splicing

Query Match 30.1%; Score 53.5; DB 2; Length 5262;
 Best Local Similarity 42.5%; Pred. No. 4.5e+02;
 Matches 17; Conservative 4; Mismatches 12; Indels 7; Gaps 2;

QY 2 GLSQEQLEHRSRSLQTLRD-----IQRLFPDEKEFTCA 35
 |||| :|| :|| :||| :||| :|||
 Db 2396 GLSQEQLEHRSRSLQTLRD-----IQRLFPDEKEFTCA 2434

RESULT 10
 B70438
 hypothetical protein aq_1596 - Aquifex aeolicus
 C:Species: Aquifex aeolicus
 C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
 C:Accession: B70438
 R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.
 V.
 Nature 392, 353-358, 1998
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A:Reference number: A70300; MUID:98196666; PMID:9537320
 A:Accession: B70438
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-227 <AQF>
 A:Cross-references: GB:AE000747; NID:g2983944; PIDN:AAC07503.1; PID:g2983955; GB:AE0
 A:Experimental source: strain VF5
 C:Genetics:
 A:Gene: aq_1596

Query Match 29.8%; Score 53; DB 2; Length 227;
 Best Local Similarity 41.4%; Pred. No. 16;
 Matches 12; Conservative 7; Mismatches 8; Indels 2; Gaps 1;

QY 5 OBOLEHRSRSLQTLRDQRM--FPDEKE 31
 :||| :||| :||| :||| :||| :|||
 Db 33 KELLEEREKLEETIRSYEEKLDSFEKEQ 61

RESULT 11
 E69957
 gamma-D-glutamyl-L-diamino acid endopeptid homolog yqgT - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C:Accession: E69957
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Be
 C:Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A:Erlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Ga
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo,
 Ketter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scan
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; S
 akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiy
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshid

A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: B69957
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-376 <KUN>
A;Cross-references: GB:299116; GB:AL009126; NID:g2634723; PIDN:CAB14414.1; PID:g2634917
A;Experimental source: strain 168
C;Genetics:
C;Superfamily: endopeptidase I

Query Match 29.8%; Score 53; DB 1; Length 376;
Best Local Similarity 52.6%; Pred. No. 28;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 9 EHRERSLQTLRDIQRMFLFP 27
DB 50 EHSKELQTLQIDIKRFLQP 68

RESULT 12
S62790
mismatch DNA recognition protein muts [validated] - *Thermus aquaticus* (fragment)
C;Species: *Thermus aquaticus*
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 26-May-2000
C;Accession: S62790
R;Takamatsu, S.; Kato, R.; Kuramitsu, S.
Nucleic Acids Res. 24, 640-647, 1996
A;Title: Mismatch DNA recognition protein from an extremely thermophilic bacterium, *Thermoplasma*
A;Reference number: S62790; MUID:96177563; PMID:8604304
A;Accession: S62790
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-818 <TAK>
A;Cross-references: EMBL:D63810
A;Note: the source is designated as *Thermus thermophilus*
C;Genetics:
A;Gene: muts
C;Function:
A;Description: implicated in DNA mismatch repair; binds to DNA and specifically recognizes
attached DNA [validated, MUID:96177563]
C;Superfamily: DNA mismatch repair protein muts

Query Match 29.8%; Score 53; DB 2; Length 818;
Best Local Similarity 42.3%; Pred. No. 67;
Matches 11; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 6 EQLEHRSLSQTLRDIQRMFLPDEKE 31
DB 465 EKVPQYRPVQTLKDRQRYTLPEMKR 490

RESULT 13
B70366
hypothetical protein aq_755 - *Aquifex aeolicus*
C;Species: *Aquifex aeolicus*
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 18-Aug-2000
C;Accession: B70366
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: B70366
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-242 <AQF>
A;Cross-references: GB:AE000706; NID:g2983327; PIDN:AA06923.1; PID:g2983334; GB:AE00065
A;Experimental source: strain VF5
C;Genetics:
A;Gene: aq_755

C;Superfamily: *Aquifex aeolicus* hypothetical protein aq_755

Query Match 29.5%; Score 52.5; DB 2; Length 242;
Best Local Similarity 45.8%; Pred. No. 20;
Matches 11; Conservative 7; Mismatches 5; Indels 1; Gaps 1;

QY 10 HRSLSQTLRDIQRMFLPDEKEF 32
DB 60 HKRTSLRKFFVREIEKMWFEAEKEF 83

RESULT 14

T24343
hypothetical protein T02B5.1 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Jun-2000
C;Accession: T24343
R;McMurray, A.
submitted to the EMBL Data Library, October 1996

A;Reference number: Z19878

A;Accession: T24343

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-705 <WIL>

A;Cross-references: EMBL:Z81112; PIDN:CAR03272.1; GSPDB:GN00023; CESP:T02B5.1

A;Experimental source: clone T02B5

C;Genetics:

A;Gene: CESP:T02B5.1

A;Map position: 5

A;Introns: 22/2; 88/2; 117/3; 185/1; 221/3; 280/3; 349/2; 427/2; 532/2; 637/1

C;Superfamily: cholinesterase; cholinesterase homology

Query Match 29.5%; Score 52.5; DB 2; Length 705;
Best Local Similarity 36.8%; Pred. No. 66;
Matches 14; Conservative 7; Mismatches 12; Indels 5; Gaps 1;

QY 3 LSEQLEH-----RERSLQTLRDIQRMFLPDEKEFTGA 35
DB 273 LSEQVENTYSCLRKSAQQLDAQLWLNQSTYFLGA 310

RESULT 15

T14802
phytochrome B - sorghum (fragment)

C;Species: *Sorghum bicolor* (sorghum)

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 05-May-2000

C;Accession: T14802

R;Childs, K.L.; Miller, F.R.; Cordonnier-Pratt, M.M.; Pratt, L.H.; Morgan, P.W.; Mull

A;Description: The Sorghum bicolor photoperiod sensitivity gene, Ma3, encodes a phyto

A;Reference number: Z18185

A;Accession: T14802

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1039 <CHI>

A;Cross-references: EMBL:U56730; NID:g1800216; PID:g1800217

A;Experimental source: cultivar 58M

C;Genetics:

A;Gene: PHYB

A;Note: Intron positions not resolved (incomplete sequence)

C;Superfamily: phytochrome; phytochrome homology

C;Keywords: chromoprotein; photoreceptor; phytochromobilin

F.233/Binding site: phytochromobilin (Cys) (covalent) #status predicted

Query Match 29.5%; Score 52.5; DB 2; Length 1039;
Best Local Similarity 51.9%; Pred. No. 1e+02;
Matches 14; Conservative 2; Mismatches 6; Indels 5; Gaps 1;

QY 3 LSEQLEHRSLSQTLRDIQRMFLPDE 29
DB 887 VSQAMLLLRDLQLIRDI-----PDE 908

Search completed: October 15, 2003, 10:33:10
Job time : 11.3074 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 4.22179 Seconds
(without alignments)
389.867 Million cell updates/sec

Title: US-09-915-543-15_COPY_349_383
Perfect score: 178
Sequence: 1 DGLSQEQLHRSRLQTLRIQRLPDEKEFTGA 35

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	178	100.0	1426	1 BCL9_HUMAN	O00512 homo sapien
2	66	37.1	1469	1 BCL9_DROME	Q961d9 drosophila
3	56	31.5	584	1 TRF4_YEAST	P53632 saccharomyc
4	55	30.9	426	1 HEM1_SYNEL	O8d153 synecococc
5	55	30.9	718	1 RHG8_HUMAN	Q9ns90 homo sapien
6	54.5	30.6	174	1 ASH3_MOUSE	Q94j17 mus musculu
7	54.5	30.6	1171	1 PHVB_ORYSA	P25764 oryza sativ
8	53	29.8	376	1 YOGT_BACSU	P54497 bacillus su
9	53	29.8	817	1 MUTS_THECA	Q921x6 thermus cal
10	53	29.8	818	1 MUTS_THETH	Q56239 thermus the
11	52.5	29.5	180	1 ASH3_HUMAN	Q9nq33 homo sapien
12	52.5	29.5	190	1 RRF_FUSNN	O8f5z9 fusobacteri
13	52.5	29.5	242	1 Y755_AQUAE	O66957 aquifex aeo
14	52.5	29.5	1178	1 PHVB_SORBI	P93527 sorghum bic
15	52	29.2	332	1 RP5B_ANASP	Q03065 anabaena sp
16	52	29.2	811	1 MUTS_THEAQ	O56215 thermus aqu
17	52	29.2	1009	1 RFA2_YEAST	Q06407 saccharomyc
18	51	28.7	237	1 YFPG_ECOLI	P37328 escherichia
19	51	28.7	899	1 POL_MMTVB	P03365 mouse mamma
20	50	28.1	425	1 RHG8_MOUSE	Q9c3p4 mus musculu
21	50	28.1	2004	1 MOZL_HUMAN	Q92794 homo sapien
22	49.5	27.8	301	1 TRT2_CHICK	P02642 gallus gall
23	49.5	27.8	830	1 DYN1_CAEEL	P39055 caenorhabdi
24	49	27.5	310	1 LDH_BACHD	Q9k5z8 bacillus ha
25	49	27.5	328	1 PLSX_STAAM	O8nx10 staphylococ
26	49	27.5	661	1 YDBJ_SCHPO	Q10369 schizosacch
27	49	27.5	1207	1 DML1_ARATH	Q9s1q6 arabidopsis
28	49	27.5	1234	1 PIP3_HUMAN	Q01970 homo sapien
29	48.5	27.2	184	1 YWCK_BACSU	P45870 bacillus su
30	48.5	27.2	1073	1 CARB_METMA	P58944 methanosarc
31	48.5	27.2	1132	1 PHVB_TOBAC	P29130 nicotiana t
32	48.5	27.2	1505	1 SCP2_RAT	O70608 rattus norv
33	48.5	27.2	1872	1 T2D1_HUMAN	P21675 homo sapien

ALIGNMENTS

RESULT 1

BCL9_HUMAN

ID BCL9_HUMAN STANDARD; PRT; 1426 AA.

AC O00512;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE B-cell lymphoma 9 protein (Bcl-9) (legless homolog).

GN BCL9.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Fetal brain;

RX MEDLINE=98158621; PubMed=9490669;

RA Willis T.G., Zalcborg I.R., Coignet L.J.A., Wlodarska I., Stul M.,

RA Jadayel D.M., Bastard C., Treleaven J.G., Catovsky D., Silva M.L.M.,

RA Dyer M.J.S.;

RT "Molecular cloning of translocation t(1;14)(q21;q32) defines a novel

gene (BCL9) at chromosome 1q21.";

RL Blood 91:1873-1881(1998).

RN [2]

RP FUNCTION.

RX MEDLINE=21952490; PubMed=11955446;

RA Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,

RA Murone M., Zuehlbig S., Basler K.;

RT "Wnt/Wingless signaling requires BCL9/legless-mediated recruitment of

pygopus to the nuclear beta-catenin-tcf complex.";

RL Cell 109:47-60(2002).

CC -!- FUNCTION: Involved in signal transduction through the wnt pathway.

CC -!- SUBUNIT: Binds to beta-catenin (CTNBL1), PYGOL1 and PYGOL2.

CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -!- TISSUE SPECIFICITY: Detected at low levels in thymus, prostate,

CC testis, ovary and small intestine, and at lower levels in spleen,

CC colon and blood.

CC -!- DISEASE: Involved in a t(1;14)(q21;q32) chromosomal translocation

CC found in a patient with precursor B-cell acute lymphoblastic

CC leukemia (ALL). This translocation leaves the coding region

CC intact, but may have pathogenic effects due to alterations in the

CC expression level of BCL9. Several cases of translocations within

CC the 3' untranslated region of BCL9 have been found in B-cell

CC malignancies.

CC -!- CAUTION: It is uncertain whether Met-1 or Met-27 is the initiator.

CC -!- CAUTION: Ref.1 sequence differs from that shown due to a

CC frameshift in position 1391.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC

EMBL; Y13620; CAA73942.1; ALT_FRAME.
DR Genew; HGNC:1008; BCL9.
DR MIM; 602597; -
GO GO:0007048; P:chromosome translocation; Proto-oncogene;
KW Nuclear protein; Chromosomal translocation; Proto-oncogene;
Wnt signaling pathway.
FT DOMAIN 231 1378 PRO-RICH.
FT DOMAIN 347 377 CTNNB1-BINDING.
FT DOMAIN 331 335 POLY-PRO 1.
FT DOMAIN 514 517 POLY-PRO 2.
FT DOMAIN 900 903 POLY-ALA.
FT DOMAIN 970 973 POLY-PRO 3.
SQ SEQUENCE 1426 AA; 149314 MW; A240A487716B7FLB CRC64;

Query Match 100.0%; Score 178; DB 1; Length 1426;
Best Local Similarity 100.0%; Pred. No. 3.1e-15;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DGLSQQLEHRSLSQTLRDIQRMFLPDEKFTGA 35
DB 349 DGLSQQLEHRSLSQTLRDIQRMFLPDEKFTGA 383

RESULT 2
BCU9_DROME STANDARD; PRT; 1469 AA.
ID BCU9_DROME AC Q961D9; Q9VAD2;
AC AC Q961D9; Q9VAD2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Bcl-9 homolog (legless protein).
GN BCL9 OR LGS OR CG2041.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX ADAMS M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RG George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RH Sutton G.C., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RI Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RJ Wan K.H., Doyle C., Baxter E.G., Helt C., Nelson C.R., Miklos G.L.G.,
RK Abell J.F., Agayani A., An H.-J., Andrews-Frankkoch C., Balowin D.,
RL Ballou R.M., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RM Beeson K.Y., Bonas P., Bowman M.R., Bouck J., Brokstein P., Brotter P.,
RN Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RO Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RU de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RV Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RW Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RX Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RY Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
SZ Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
TA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
TB Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
TC Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
TD Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
TE Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,
TF Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
TG Mount S.M., Moy M., Murphy B., Murphy L., Muzy D.M., Nelson D.L.,
TH Nelson D.R., Nelson K.A., Nixon K.A., Nusskern D.R., Pacleeb J.M.,
TI Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
TJ Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
TK Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
TL Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
TM Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
TN Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,


```

CC -1- SUBUNIT: Homodimer.
CC -1- PPM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.
CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) dimerization domains.
CC -1- SIMILARITY: CONTAINS 1 histidine kinase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X57563; CAA40795.2; -.
CC PIR; S14065; S14065.
CC Gramene; P25764; -.
CC InterPro; IPR003594; ATPbind_ATPase.
CC InterPro; IPR003018; GAF.
CC InterPro; IPR003661; His_kinA.
CC InterPro; IPR005467; His_kinase.
CC InterPro; IPR000014; PAS_domain.
CC InterPro; IPR001294; Phytochrome.
CC Pfam; PF01590; GAF; 1.
CC Pfam; PF02518; HATPase_c; 1.
CC Pfam; PF00512; HiskA; 1.
CC Pfam; PF00989; PAS; 2.
CC Pfam; PF00360; phytochrome; 1.
CC PRINTS; PR01033; PHYTOCHROME.
CC SMART; SM00065; GAF; 1.
CC SMART; SM00387; HATPase_c; 1.
CC SMART; SM00388; HiskA; 1.
CC SMART; SM00091; PAS; 2.
CC TIGRFAMs; TIGR00229; sensory_box; 2.
CC PROSITE; PS50109; HIS_KIN; 1.
CC PROSITE; PS50112; PAS; 2.
CC PROSITE; PS00245; PHYTOCHROME_1; 1.
CC PROSITE; PS50046; PHYTOCHROME_2; 1.
CC Transcription regulation; Photoreceptor; Phytochrome; Chromophore;
KW Repeat; Multigene family.
KW DOMAIN 561 732 PAS 1.
FT DOMAIN 795 866 PAS 2.
FT DOMAIN 943 1161 HISTIDINE KINASE.
FT DOMAIN 39 51 POLY-GLY.
FT BINDING 364 364 CHROMOPHORE (BY SIMILARITY).
SQ SEQUENCE 1171 AA; 128384 MW; E8292E88B769BF16 CRC64;

Query Match 30.6%; Score 54.5; DB 1; Length 1171;
Best Local Similarity 48.1%; Pred. NO. 30;
Matches 13; Conservative 4; Mismatches 5; Indels 5; Gaps 1;

QY 3 LSQQLERHRSLSQTLRDIQRLPDE 29
DB 1019 VSQVMIQLRDLQLRDI-----PDE 1040

RESULT 8
YQGT_BACSU STANDARD; PRT; 376 AA.
AC P54497;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yqgt.
GN YQGT.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=97124195; PubMed=8969508;
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,

```

```

RA Kobayashi Y.;
RT *Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
RT the bacillus subtilis genome containing the skin element and many
RT sporulation genes.*;
RL Microbiology 142:3103-3111(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guilseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klarr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic C.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT *The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.*;
RL Nature 390:249-256(1997).
CC -1- SIMILARITY: STRONG, TO B.SPHERICUS GAMMA-D-GLUTAMYL-L-DIAMINO
CC ACID ENDOPEPTIDASE I (EC 3.4.99.11).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D84432; BAA12523.1; -.
CC EMBL; Z99116; CAB14414.1; -.
CC PIR; E69957; E69957.
CC MEROPS; M14.008; -.
CC Subtilist; BG11687; yqgt.
CC InterPro; IPR000834; Zn_carbopept.
CC Pfam; PF00246; Zn_carbopept; 1.
CC PRINTS; PR00765; CRBOXYPTASEA.
CC SMART; SM00631; Zn_pept; 1.
CC Hypothetical protein; Hydrolase; Metalloprotease; Zinc;
KW Complete proteome.
SQ SEQUENCE 376 AA; 43439 MW; 1075CDC124823BB4 CRC64;

Query Match 29.8%; Score 53; DB 1; Length 376;
Best Local Similarity 52.6%; Pred. NO. 13;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 9 EHRERSLQTLRDIQRLPFP 27
DB 50 EHSKELTIQIDKRFLOP 68

```



```

GN ASCL3 OR SGN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=21643927; PubMed=11784080;
RA Yoshida S., Ohno K., Takakura A., Takebayashi H., Okada T., Abe K.,
RA Nabeshima Y.;
RT "Sgn1, a basic helix-loop-helix transcription factor delineates the
RT salivary gland duct cell lineage in mice.";
RL Dev. Biol. 240:517-530(2001).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=21418998; PubMed=11528127;
RA Anid C., Bahr A., Mujica A., Sampson N., Bikar S.E., Winterpacht A.,
RA Zabel B., Hankeln T., Schmidt E.R.;
RT "Comparative genomic sequencing reveals a strikingly similar
RT architecture of a conserved syntenic region on human chromosome
RT 11p15.3 (including gene SPS) and mouse chromosome 7.";
RL Cytogenet. Cell Genet. 93:284-290(2001).
CC -1- FUNCTION: Transcriptional repressor. Inhibits myogenesis (By
CC similarity).
CC -1- SUBUNIT: Efficient DNA binding requires dimerization with another
CC bHLH protein (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. ASC SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AB046450; BAB83913.1; -.
DR Genbank; AF0001092; ASCL3.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00038; HLH_1; 1.
DR PROSITE; PS00888; HLH_2; 1.
KW Transcription regulation; Repressor; DNA-binding; Nuclear protein.
FT DNA_BIND 95 105
FT DOMAIN 106 145
FT CONFLICT 54 54 R -> L (IN REF. 1).
SQ SEQUENCE 180 AA; 20796 MW; C9B90C885DE94F81 CRC64;
Query Match 29.5%; Score 52.5; DB 1; Length 180;
Best Local Similarity 36.1%; Pred. No. 6.6;
Matches 13; Conservative 7; Mismatches 9; Indels 7; Gaps 1;
QY 3 LSOQLEHRSLSQTLRD-----IQRLFPDEKE 31
DB 119 LPEYLEKRLSKVETLRAAKYINYLQSLYPDEAE 154
-----
RESULT 12
RF_FUSNN STANDARD; PRT; 190 AA.
AC Q8529;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ribosome recycling factor (Ribosome releasing factor) (RRF).
GN FRR OR FNL623.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.

```

```

OX NCBI_TaxID=76856;
RN [1]
RN SEQUENCE FROM N.A.
RX STRAIN=ATCC 25386;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyrpides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25386.";
RL J. Bacteriol. 184:2005-2018(2002).
CC -1- FUNCTION: Responsible for the release of ribosomes from messenger
CC RNA at the termination of protein biosynthesis. May increase the
CC efficiency of translation by recycling ribosomes from one round of
CC translation to another (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE RRF FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AE010470; AAL93738.1; -.
DR HAMAP; MF_00040; -.
DR InterPro; IPR002661; RRF.
DR Pfam; PF01765; RRF; 1.
DR ProDom; PD004103; RRF; 1.
DR TIGRFAMs; TIGR00496; rrf; 1.
KW Protein biosynthesis; Complete proteome.
SQ SEQUENCE 190 AA; 21438 MW; 81BD4A7A293B2814 CRC64;
Query Match 29.5%; Score 52.5; DB 1; Length 190;
Best Local Similarity 36.1%; Pred. No. 7;
Matches 13; Conservative 8; Mismatches 10; Indels 5; Gaps 1;
QY 3 LSOQLEHRSLSQTLRD-----IQRLFPDEKEFT 33
DB 153 ISEDLAKKEETNQTLDKYVKEIDDLAKKEKIT 188
-----
RESULT 13
Y755_AQAE STANDARD; PRT; 242 AA.
AC O66957;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_755.
GN AQ_755.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RN SEQUENCE FROM N.A.
RX STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aulay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```



```
RT Cyanobacterium Anabaena sp. strain PCC 7120.*;
RL DNA Res. 8:205-213(2001).
CC -!- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
CC THEN IS RELEASED.
CC -!- SIMILARITY: Belongs to the sigma-70 factor family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M95760; AAA22046.1; -.
DR EMBL; AP003602; BAE77258.1; -.
DR PIR; AD2541; AD2541.
DR HSP; P00579; 1SIG.
DR InterPro; IPR000943; Sigma_70.
DR Pfam; PF00140; sigma70_r1_2; 1.
DR Pfam; PF04542; sigma70_r2; 1.
DR Pfam; PF04539; sigma70_r3; 1.
DR Pfam; PF04545; sigma70_r4; 1.
DR PRINTS; PR00046; SIGMA70FCT.
DR PROSITE; PS00715; SIGMA70.1; 1.
DR PROSITE; PS00716; SIGMA70.2; 1.
DR Transcription regulation; Sigma factor; DNA-directed RNA polymerase;
KW DNA-binding; Plasmid; Complete proteome.
FT DOMAIN 125 138 POLYMERASE CORE BINDING (POTENTIAL).
FT DNA_BIND 294 313 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 332 AA; 38431 MW; 81323E670CCDF46A CRC64;

Query Match 29.2%; Score 52; DB 1; Length 332;
Best Local Similarity 35.5%; Pred. No. 15;
Matches 11; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

Oy 1 DGLSQOLEHRESRSLQTLRDIQRMFLPDEKE 31
Db ||:| |: || |:|:|:| |:|:|:|
248 DCMSPERYARELLYQDIHNLAKLTPOQKE 278
```

Search completed: October 15, 2003, 10:31:02
Job time : 6.22179 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:33 ; Search time 20.2918 Seconds
(without alignments)
445.097 Million cell updates/sec

Title: US-09-915-543-15_COPY_349_383
Perfect score: 178
Sequence: 1 DGLSQLEHRESLQTLRDIQRMLFPDEKEFTGA 35

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:.*
1: sp_archaea:.*
2: sp_bacteria:.*
3: sp_fungi:.*
4: sp_human:.*
5: sp_invertebrate:.*
6: sp_mammal:.*
7: sp_mhc:.*
8: sp_organelle:.*
9: sp_phase:.*
10: sp_plant:.*
11: sp_rodent:.*
12: sp_virus:.*
13: sp_vertebrate:.*
14: sp_unclassified:.*
15: sp_virus:.*
16: sp_bacteriaph:.*
17: sp_archaeap:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	57	32.0	1034	16 Q8Z8Y6	Q8Z8Y6 salmonella
2	57	32.0	1046	16 Q8ZRE3	Q8ZRE3 salmonella
3	56	31.5	859	4 Q9NT51	Q9nt51 homo sapien
4	56	31.5	1132	4 Q9Y2H2	Q9y2h2 homo sapien
5	56	31.5	1208	5 Q9VXU1	Q9vxu1 drosophila
6	56	31.5	1398	5 Q9VXU2	Q9vxu2 drosophila
7	55	30.9	295	16 Q8UCA0	Q8uca0 agrobacteri
8	55	30.9	426	16 Q8DI53	Q8di53 synechococc
9	55	30.9	433	4 Q8IZM6	Q8izm6 homo sapien
10	55	30.9	818	17 Q8TGZ1	Q8tgz1 methanopyru
11	53.5	30.1	155	11 Q8VD56	Q8vd56 rattus norv
12	53.5	30.1	4957	4 O14687	O14687 homo sapien
13	53.5	30.1	5262	4 O14686	O14686 homo sapien
14	53	29.8	227	16 Q67532	Q67532 aquifex aeo
15	53	29.8	411	16 Q8PRD9	Q8prd9 xanthomonas
16	52.5	29.5	673	5 O01302	O01302 caenorhabdi

17	52	29.2	399	10 Q8WOM3	Q8wom3 oryza sativ
18	52	29.2	572	16 Q9HTQ7	Q9htq7 pseudomonas
19	52	29.2	616	12 P87708	P87708 fowlpox vir
20	52	29.2	946	4 Q96OE4	Q96qe4 homo sapien
21	52	29.2	1009	3 Q06407	Q06407 saccharomyc
22	52	29.2	1786	12 Q9J599	Q9j599 fowlpox vir
23	52	29.2	2607	11 Q8BT18	Q8bt18 mus musculu
24	51.5	28.9	1146	10 Q9FPQ2	Q9fpq3 populus tri
25	51.5	28.9	1151	10 Q9FPQ3	Q9fpq3 populus tri
26	51	28.7	102	16 Q82FD6	Q8zfd6 yersinia pe
27	51	28.7	108	16 Q8D096	Q8d096 yersinia pe
28	51	28.7	239	16 Q8FF28	Q8ff28 escherichia
29	51	28.7	258	11 Q8BVP0	Q8bvp0 mus musculu
30	51	28.7	329	10 Q8LG66	Q8lge6 arabidopsis
31	51	28.7	329	10 Q9C975	Q9c975 arabidopsis
32	51	28.7	337	5 Q62022	Q62022 caenorhabdi
33	51	28.7	341	4 Q8IY13	Q8iy13 homo sapien
34	51	28.7	449	5 Q8IOM4	Q8iom4 caenorhabdi
35	51	28.7	848	16 Q8FMB8	Q8fmb8 corynebacte
36	51	28.7	1047	16 Q8XEJ6	Q8xej6 escherichia
37	51	28.7	1464	5 Q23995	Q23995 drosophila
38	51	28.7	1464	5 Q9VC47	Q9vc47 drosophila
39	51	28.7	1464	5 Q24132	Q24132 drosophila
40	51	28.7	1755	15 Q56220	Q56220 mouse mamma
41	51	28.7	1755	15 Q9I2T3	Q9i2t3 exogenous m
42	50.5	28.4	96	5 Q9VN78	Q9vn78 drosophila
43	50.5	28.4	288	13 Q8UVZ1	Q8uvz1 brachydanio
44	50.5	28.4	657	10 Q9FNQ6	Q9fnq6 arabidopsis
45	50.5	28.4	675	4 Q9HD27	Q9hd27 homo sapien

ALIGNMENTS

RESULT 1

Q8Z8Y6 PRELIMINARY; PRT; 1034 AA.
ID Q8Z8Y6
AC Q8Z8Y6;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Exonuclease Sbcc.
GN STY0429.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_taxid=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT *Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.*;
RL Nature 413:848-852(2001).
DR EMBL: AL627266; CAD08850.1; -;
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR004592; Sbcc.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02463; SMC_N; 1.
DR TIGRFAMs; TIGR00618; sbcc; 1.
KW Complete proteome.
SQ SEQUENCE 1034 AA; 116759 MW; 2513B7573626960A CRC64;

Query Match 32.0%; Score 57; DB 16; Length 1034;
Best Local Similarity 42.9%; Pred. No. 68;

```

Matches 12; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 3 LSQEQLEHRSLSQTLRDIOQLMFLPDEK 30
   |:|:|:| | | | | | | | | | | | | |
Db 213 LADEQLQQLASLQALTDDEKRLADQ 240

RESULT 2
Q8ZRE3
ID Q8ZRE3 PRELIMINARY; PRT; 1046 AA.
AC Q8ZRE3;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE ATP-dependent dsDNA exonuclease.
GN SBCC OR STM0395.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC SFRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT *Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.*;
RL Nature 413:852-856(2001).
DR EMBL: AE008713; AAL19349.1; -.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR004592; Sbcc.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02463; SMC_N; 1.
DR TIGRfams: TIGR00618; sbcc; 1.
KW Exonuclease; Complete proteome.
SQ SEQUENCE 1046 AA; 117824 MW; BA565CA3BDAD0C82 CRC64;

Query Match 32.0%; Score 57; DB 16; Length 1046;
Best Local Similarity 42.9%; Pred. No. 69;
Matches 12; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 3 LSQEQLEHRSLSQTLRDIOQLMFLPDEK 30
   |:|:|:| | | | | | | | | | | | | |
Db 213 LADEQLQQLASLQALTDDEKRLADQ 240

RESULT 3
Q9NT51
ID Q9NT51 PRELIMINARY; PRT; 859 AA.
AC Q9NT51;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFZP434P1818.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Testis;
RA Ottenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL137528; CAB70792.1; -.
DR InterPro: IPR002013; Syja_N.
DR Pfam: PF02383; Syja_N; 1.
DR PROSITE: PS50275; SAC; 1.
KW Hypothetical protein.

```

```

FT NON_TER 1
SQ SEQUENCE 859 AA; 96781 MW; BCBB47C8B726D76 CRC64;

Query Match 31.5%; Score 56; DB 4; Length 859;
Best Local Similarity 33.3%; Pred. No. 76;
Matches 10; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 5 QEQLEHRSLSQTLRDIOQLMFLPDEKFTG 34
   |:|:|:| | | | | | | | | | | | | |
Db 317 ENQRSHQELISQLQSYMKLLLPDDEKFG 346

RESULT 4
Q9Y2H2
ID Q9Y2H2 PRELIMINARY; PRT; 1132 AA.
AC Q9Y2H2;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Hypothetical protein KIAA0966.
GN KIAA0966.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Brain;
RX MEDLINE=99246033; PubMed=10231032;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT *Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.*;
RL DNA Res. 6:63-70(1999).
DR EMBL: AB023183; BAA76810.1; -.
DR InterPro: IPR002013; Syja_N.
DR Pfam: PF02383; Syja_N; 1.
DR PROSITE: PS50275; SAC; 1.
KW Hypothetical protein.
SQ SEQUENCE 1132 AA; 128407 MW; 853719FC0AD455CD CRC64;

Query Match 31.5%; Score 56; DB 4; Length 1132;
Best Local Similarity 33.3%; Pred. No. 1e+02;
Matches 10; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 5 QEQLEHRSLSQTLRDIOQLMFLPDEKFTG 34
   |:|:|:| | | | | | | | | | | | | |
Db 590 ENQRSHQELISQLQSYMKLLLPDDEKFG 619

RESULT 5
Q9VXU1
ID Q9VXU1 PRELIMINARY; PRT; 1208 AA.
AC Q9VXU1;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE CG32587 protein.
GN CG32587 OR CG6350 OR CG7821.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

```



```

Db 134 GFSQTLLEKRLRLHLRLDRV 157

RESULT 11
Q8VD56 PRELIMINARY; PRT; 155 AA.
AC Q8VD56;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE BCH domain-containing Cdc42GAP-like protein.
GN BPGAP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shang X., Zhou Y.T., Low B.C.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF544240; AAM0769.1; -; OAE4B42A04AE1D3 CRC64;
SQ SEQUENCE 433 AA; 49691 MW; 0AE4B42A04AE1D3 CRC64;

Query Match 30.9%; Score 55; DB 4; Length 433;
Best Local Similarity 52.2%; Pred. No. 51;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQQLHRSLSQTLRLDIQRM 24
|| || | | | | | | | | | | | | | | | |
Db 224 GLRTEGLFRSSASVQTVREIQL 246

RESULT 10
Q8TG21 PRELIMINARY; PRT; 818 AA.
AC Q8TG21;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Archaea-specific superfamily II helicase.
GN MK0835.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezheva V.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozyavkin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
and monophyly of archaeal methanogens.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
DR EMBL: AF010374; AAM02048.1; -;
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00270; DEAD; 1.
DR Pfam: PF00271; helicase_C; 1.
DR SMART: SM00382; AAA; 1.
DR SMART: SM00487; DEAD; 1.
DR SMART: SM00490; HELIC; 1.
KW Helicase; Complete proteome.
SQ SEQUENCE 818 AA; 91715 MW; C2136200A710817E CRC64;

Query Match 30.9%; Score 55; DB 17; Length 818;
Best Local Similarity 54.2%; Pred. No. 98;
Matches 13; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 2 GLSQQLHRSLSQTLRLDIQRM 25
|| || | | | | | | | | | | | | | | | |

```

```

Db 134 GFSQTLLEKRLRLHLRLDRV 157

RESULT 11
Q8VD56 PRELIMINARY; PRT; 155 AA.
AC Q8VD56;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Putative BHLH transcription factor (Fragment).
GN SGN-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Sprague-Dawley;
RX MEDLINE=21643927; PubMed=11784080;
RA Yoshida S., Ohbo K., Takakura A., Takebayashi H., Okada T., Abe K.,
RA Nabeshima Y.;
RT "Sgn1, a Basic Helix-Loop-Helix Transcription Factor Delineates the
RT Salivary Gland Duct Cell Lineage in Mice.";
RL Dev. Biol. 240:517-530(2001).
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
CC EMBL: AB046449; BAB83912.1; -;
DR InterPro: IPR001092; HLH_Basic.
DR Pfam: PF00010; HLH; 1.
DR SMART: SM00353; HLH; 1.
DR PROSITE: PS00038; HLH_1; 1.
DR PROSITE: PS00888; HLH_2; 1.
DR NON_TER 1
FT NON_TER 1
SQ SEQUENCE 155 AA; 18146 MW; 6E8ABD1DA16B88D6 CRC64;

Query Match 30.1%; Score 53.5; DB 11; Length 155;
Best Local Similarity 36.1%; Pred. No. 28;
Matches 13; Conservative 7; Mismatches 9; Indels 7; Gaps 1;

QY 3 LSQQLHRSLSQTLRLDIQRM 31
| : | | | | : | | | | : | | | |
Db 100 LPEDYLEKRLSKVETLRAAKYISYLSQSLYPDETE 135

RESULT 12
O14687 PRELIMINARY; PRT; 4957 AA.
AC O14687;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE ALR.
GN ALR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97388474; PubMed=9247308;
RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
RA Canaan E.;
RT "Structure and expression pattern of human ALR, a novel gene with
RT strong homology to ALL-1 involved in acute leukemia and to Drosophila
RT trithorax.";
RL Oncogene 15:549-560(1997).
CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
DR EMBL: AF010404; AAC51735.1; -;
DR InterPro: IPR003889; FYrich_C.
DR InterPro: IPR003888; FYrich_N.
DR InterPro: IPR000910; HMG_12_box.

```

DR InterPro: IPR003616; PostSET.
DR InterPro: IPR006118; Recombinase.
DR InterPro: IPR001214; SET.
DR InterPro: IPR001965; Znf_PHD.
DR Pfam: PF00628; PHD; 3.
DR Pfam: PF00856; SET; 1.
DR SMART: SM00542; FYRC; 1.
DR SMART: SM00541; FYRN; 1.
DR SMART: SM00398; HMG; 1.
DR SMART: SM00249; PHD; 4.
DR SMART: SM00508; PostSET; 1.
DR SMART: SM00317; SET; 1.
DR PROSITE; PS00398; RECOMBINASES_2; 1.
DR PROSITE; PS0280; SET; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
SQ SEQUENCE 4957 AA; 531840 MW; 1026562E1419CE8D CRC64;

Query Match 30.1%; Score 53.5; DB 4; Length 4957;
Best Local Similarity 42.5%; Pred. No. 9.6e+02;
Matches 1; Conservative 4; Mismatches 12; Indels 7; Gaps 2;

Qy 2 GLSQEQLEHRSRSLQTLRD-----IQRMFPDEKEFTGA 35
||||:||||:||||:
Db 2091 GLSQTELE-KORQQRLELLIRQIQIRNLTLRQEKETAAA 2129
||||:||||:||||:

RESULT 13
O14686
ID O14686 PRELIMINARY; PRU: 5262 AA.
AC O14686;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE ALR.
GN ALR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
FP [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=97388474; PubMed=9247308;
RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
RA Canaan E.;
RT "Structure and expression pattern of human ALR, a novel gene with
RT strong homology to ALL-1 involved in acute leukemia and to Drosophila
RT trithorax";
RL Oncogene 15:549-560(1997).
CC -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
DR EMBL: AF010403; AACSI734.1; -.
DR Genbank: HGNC:7133; MLL2.
DR InterPro: IPR003889; FYrich.C.
DR InterPro: IPR003888; FYrich.N.
DR InterPro: IPR000910; HMG_12_box.
DR InterPro: IPR003616; PostSET.
DR InterPro: IPR006118; Recombinase.
DR InterPro: IPR001214; SET.
DR InterPro: IPR001965; Znf_PHD.
DR InterPro: IPR001841; Znf_ring.
DR Pfam: PF00628; PHD; 5.
DR Pfam: PF00856; SET; 1.
DR SMART: SM00542; FYRC; 1.
DR SMART: SM00541; FYRN; 1.
DR SMART: SM00398; HMG; 1.
DR SMART: SM00249; PHD; 7.
DR SMART: SM00508; PostSET; 1.
DR SMART: SM00184; RING; 3.
DR SMART: SM00317; SET; 1.
DR PROSITE; PS00398; RECOMBINASES_2; 1.
DR PROSITE; PS50280; SET; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.

RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira R.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT *Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities";
RL Nature 417:459-463(2002).
DR EMBL; AE011625; AAM34916.1; -.
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01551; Peptidase_M37; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 411 AA; 44337 MW; E193496126355426 CRC64;

Query Match 29.8%; Score 53; DB 16; Length 411;
Best Local Similarity 36.4%; Pred. No. 89;
Matches 12; Conservative 4; Mismatches 17; Indels 0; Gaps 0;

QY 2 GLSQEQLEHRRSLQTLRDIQRMLEFPDKEFTG 34
| | : | | | | : | | |
Db 17 GASAQSQRETERKQLQRLDELKLTISADRRLEGG 49

Search completed: October 15, 2003, 10:35:47
Job time : 23.2918 secs